Testing for Causality with Wald Tests under Nonregular Conditions

D I S S E R T A T I O N

zur Erlangung des akademischen Grades
doctor rerum politicarum
(dr. rer. pol.)
im Fach Ökonometrie

eingereicht an der
Wirtschaftswissenschaftlichen Fakultät
Humboldt-Universität zu Berlin

von
Frau Dipl.-Volkswirtin Maike M. Burda
geboren am 21.02.1966 in Neumünster

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jürgen Mlynek

Dekan der Wirtschaftswissenschaftlichen Fakultät:
Prof. Dr. Lutz Hildebrandt

Gutachter:

1. Prof. Dr. Helmut Lütkepohl
2. Prof. Dr. Wolfgang Härdle

eingereicht am:	28. August 2001
Tag der mündlichen Prüfung:	4. Dezember 2001

Abstract

The concepts of standard Granger causality and impulse response analysis are often used to investigate causal relationships between variables in vector autoregressive (VAR) models. In VAR models with more than two variables, the concept of standard Granger causality can be extended by studying prediction improvement at forecast horizons greater than one. The causal relationships which arise under this extended Granger causality concept are compared to those arising under the standard Granger causality concept (one–step forecasts) and those arising with impulse–response–analysis. In particular, it is illustrated inhowfar the extended Granger causality concept can be understood as a generalization of the standard Granger causality concept and even of impulse–response–analysis.
If causality is measured at forecast horizons greater than one, and if there are more than two variables in the VAR system, the null hypothesis that one variable is not causal for another variable implies restrictions which are a nonlinear function of the VAR coefficients. (In nonstationary VAR models, nonlinear restrictions already arise under the standard Granger causality concept.) Due to the special form of the restrictions, the standard Wald test may no longer have the usual asymptotic chisquare–distribution under the null hypothesis. This problem is commonly neglected in practice. However, Example 4.1, Corollary 4.1 and Proposition 4.1 of this thesis illustrate that this problem is not irrelevant. Furthermore, Propositions 5.1 and 5.2 show that this problem may be overcome, at least in stationary VAR models, by using either a randomized Wald test or a Wald test with generalized inverse. Size and Power of these modified Wald tests relative to the standard Wald test are investigated in a small simulation study for different stationary, trivariate VAR(1) models. Moreover, the pros and cons of alternative testing strategies (bootstrap, sequential tests) are summarized in a brief overview.

Keywords:
Granger causality, impulse response analysis, vectorautoregressive models, multi–step forecasts, Wald tests

Zusammenfassung

Das Kausalitätskonzept von Granger und die Impuls–Antwort–Analyse sind zwei Konzepte, die häufig verwendet werden, um kausale Beziehungen zwischen zwei Variablen in vektorautoregressiven (VAR) Modellen zu untersuchen. Wenn das VAR Modell mehr als zwei Variablen umfasst, besteht eine Erweiterung des Standard Granger Kausalitätskonzepts darin, Kausalität an höheren Prognosehorizonten zu messen. Die Kausalitätsbeziehungen unter diesem erweiterten Granger Kausalitätskonzept werden mit denen bei Standard Granger Kausalität (Ein–Schritt–Prognose) und mit Kausalität im Sinne der Impuls–Antwort–Analyse verglichen. Es wird insbesondere dargestellt, inwiefern das erweiterte Granger Kausalitätskonzept als Verallgemeinerung der letztgenannten Konzepte aufgefasst werden kann.
Wenn Kausalität an Prognosehorizonten grösser als eins gemessen wird und das VAR Modell mehr als zwei Variablen umfasst, impliziert die Nullhypothese, dass eine Variable nicht kausal für eine andere Variable sei, nichtlineare Restriktionen auf die VAR Koeffizienten. (In nichtstationären VAR Modellen treten nichtlineare Restriktionen sogar schon unter dem Standard Granger Kausalitätskonzept auf.) Aufgrund der speziellen Form der Restriktionen kann es vorkommen, dass die Standard Wald Statistik nicht mehr die übliche asymptotische Chiquadrat–Verteilung hat. Dieses Problem wird im allgemeinen in der Praxis ignoriert. Beispiel 4.1, Proposition 4.1 und Korollar 4.1 zeigen jedoch, dass dieses Problem nicht irrelevant ist. Zwei Lösungen werden in Proposition 5.1 und Proposition 5.2 in Form eines randomisierten Wald Tests sowie eines Wald Tests mit verallgemeinerter Inverse angeboten. In einer anschliessenden kleinen Simulationsstudie werden Grösse und Macht dieser modifizierten Wald Tests relativ zu der des Standard Wald Tests untersucht für verschiedene stationäre trivariate VAR(1)–Modelle. In einem kurzen Überblick werden zudem Vor– und Nachteile alternativer Testverfahren (Bootstrap, sequentielle Tests) zusammengefasst.

Schlagwörter:
Granger Kausalität, Impuls–Antwort–Analyse, Vektorautoregressive Modelle, Mehr–Schritt–Prognosen, Wald Tests

Acknowledgements

I would like to thank the many people who supported me in writing this thesis. First of all, I wish to thank my supervisor, Prof. Dr. Helmut Lütkepohl, for his guidance throughout this thesis project, in particular for his very helpful comments on preliminary versions, his steady encouragement to finish this thesis, and his patience.
I am also indebted to Prof. Dr. Wolfgang Härdle for helpful suggestions on bootstrapping and for making available the software XploRe. I would like to thank both professors for creating a positive working atmosphere which encouraged collaboration and stimulating discussions with collegues and guest researchers.
My thanks also go to my former collegues, in particular to Dr. Jörg Breitung, Dr. Marlene Müller and Alexander Benkwitz for helpful suggestions, and to Dr. Sigbert Klinke for helping to install various software.

I am also grateful to the Deutsche Forschungsgemeinschaft (DFG) and the Sonderforschungsbereich SFB 373 ("Quantification and Simulation of Economic Processes") of the Humboldt–University, Berlin, for their financial support and the chance to present my research at various workshops and seminars.

Finally, I owe special thanks to my husband Michael for inspiring conversations and great support of various forms, to Maxwell, Robert and Henry who taught me that working time is a scarce good and has to be used efficiently, and to my parents for giving me the chance to do this dissertation. Last but not least, I would like to thank all my friends who encouraged me throughout the last years, in particular Ines, Sabine, Anja, Jeannette & Mario and Judy.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In every economy, economic agents would like to know about the likely future evolution of and the dynamic interrelationships between economic variables. These questions can be studied in a multiple time series context. The joint data generating process of a set of multiple time series can often be modeled as or approximated by a linear vector autoregressive (VAR) process. Let y_t denote a k–dimensional vector of variables $y_{i,t}$, $i = 1, \ldots, k$. In a linear VAR model, every variable $y_{i,t}$ is allowed to depend linearly on its own history $y_{i,t-1},\ y_{i,t-2}, \ldots$ as well as on past values of the other variables in the VAR sytem $y_{j,t-1},\ y_{j,t-2}, \ldots, j \neq i$, on deterministic variables, e.g. a constant or a deterministic trend term, and on an error term. The error term is usually assumed to have zero mean and constant variance, and to be uncorrelated over time.

Apart from the restrictive assumption of linear dependence, VAR models allow for a fairly general functional form: every variable is interrelated with every other variable. It is left open to the structural analysis to specify the functional form by ruling out all those relationships between variables which are not supported by the data.

Two instruments to study the dynamic structure of a VAR system are the analysis of Granger causality and impulse response analysis. In VAR models, zero responses of one variable $y_{i,t}$ to a one–time, one–unit shock in another variable $y_{j,t}$ can be modeled as nonlinear restrictions on the vector autoregressive coefficients. Similar nonlinear restrictions arise with a test of Granger causality at higher forecast horizons $h > 1$. Estimation of the vector autoregressive coefficients under nonlinear restrictions can be tedious. Consequently Wald tests, which do not require estimation under the null

hypothesis, are often preferred to Lagrange Multiplier and Likelihood Ratio Tests. Under certain regularity conditions, the standard Wald statistic has an asymptotic χ^2–distribution (see e.g. Engle (1984), Buse (1982)). Nonlinear restrictions may violate these regularity conditions. In this case, inference based on the standard Wald test may be misleading (see e.g. Andrews (1987), Lütkepohl (1993), Boudjellaba et al. (1992a, 1992b), Dufour & Renault (1994, 1998) and Lütkepohl & Breitung (1997)). Although this problem is well–known, it is often ignored, for instance in tests of zero impulse response coefficients (see e.g. Lütkepohl & Breitung (1997)). The main concern of this thesis is to show that the problem is not irrelevant, and can be taken care of by using alternative Wald statistics, which continue to have an asymptotic χ^2–distribution even when the regularity conditions are not fulfilled.

The class of models considered in this thesis are stationary and nonstationary linear vector autoregressive processes with two or more variables. The VAR model as well as alternative representations, for example the moving average representation, the error correction representation or the common trends representation, are covered in Chapter 2.

In Chapter 3, the concepts of Granger causality at forecast horizons $h \geq 1$ and impulse response analysis are presented:
Based on work by Wiener (1956), Granger (1969) defines a variable y_1 to be causal for another variable y_2, if the information in $y_{1,t},\ y_{1,t-1}, \ldots$ helps to improve the prediction of $y_{2,t+1}$. If the information in the past and present of $y_{1,t}$ does not help to improve the one–step ahead forecast of $y_{2,t}$, then y_1 is called Granger noncausal for y_2. The fundamental idea of this definition of causality is that *past and present may cause the future but the future cannot cause the past* (Granger (1980, Axiom A)).
Impulse response analysis also rests on the idea that the cause precedes the effect in time. However, a causal relationship is now interpreted as a stimulus–response–mechanism: if y_1 is causal for y_2, variations in y_1 should stimulate a response of y_2. The stimulus is modeled as a one–time exogenous impulse in $y_{1,t}$. If this impulse changes the forecasts of $y_{2,t+h}$ for at least one forecast horizon $h \geq 1$, then y_1 may be

called causal for y_2. If the responses of $y_{2,t+h}$ are zero for all forecast horizons $h \geq 1$, y_1 may be called noncausal for y_2.

By restricting prediction to a forecast horizon of one period, the original concept of Granger just considers direct flows of information from one variable to another. However, in VAR models with three or more variables, information in variable $y_{1,t}$, $y_{1,t-1}, \ldots$ can also be passed on to variable $y_{2,t+h}$ through the other variables $y_{3,t}$, $y_{3,t-1}, \ldots$, $y_{4,t}$, $y_{4,t-1}$, ... The higher the number of variables in the VAR system and the higher the lag order, the higher the number of combinations via which information from y_1 may run to y_2. Dufour & Renault (1994, 1998) have shown that impulse response analysis takes into account some but not all of these combinations. On the other hand, investigating Granger causality at all forecast horizons $h \geq 1$ ensures that all possible combinations are analyzed.

Illustration of this extended concept of Granger causality and comparison with impulse response analysis are the main concerns of Chapter 3. The first part of Chapter 3 considers the type of causality that is consistent with standard and extended Granger causality and impulse response analysis. In the second part of Chapter 3, formal definitions of Granger causality and causality in terms of impulse response analysis are given, and the restrictions for Granger noncausality at forecast horizons $h \geq 1$ and for noncausality in terms of impulse response analysis are explored. Note, that different restrictions arise with different representations of the VAR model. Examples help to illustrate the restrictions as well as the relationships between causality in the sense of Granger (1969) and causality in terms of impulse response analysis.

Chapters 4 and 5 deal with testing the restrictions which have been explored in the preceding chapter. In stationary VAR models, Granger noncausality at forecast horizons $h > 1$ as well as noncausality in terms of impulse response analysis hold under nonlinear restrictions on the vector autoregressive coefficients. Chapter 4 illustrates that these nonlinear restrictions may violate the regularity condition of the Wald statistic. As consequence, the Wald statistic may no longer have an asymptotic χ^2–distribution. This problem is not limited to tests of Granger noncausality at higher forecast horizons or impulse response analysis, but can also arise in general with a certain type of

nonlinear restrictions (Gaffke et al. (1999)).

In nonstationary VAR models, even linear restrictions may violate the regularity condition of the Wald statistic if the cointegration space is restricted under the null hypothesis. The standard Granger causality test is only one example. The intensive work on integration and cointegration of the last two decades has already produced some solutions for standard Granger causality tests in nonstationary VAR models. A review of these suggestions will be given in Chapter 4. However, the problem that nonlinear restrictions may violate the regularity condition of the standard Wald statistic has not received much attention in the literature (see however Boudjellaba et al. (1992a, 1992b), Lütkepohl & Burda (1997) and Gaffke et al. (1999) for solutions). The examples of Chapter 4 illustrate the relevance of this problem for extended Granger causality tests.

Chapter 5 then presents alternative Wald statistics which all have an asymptotic χ^2–distribution under the null hypothesis of Granger noncausality at all forecast horizons $h \geq 1$, also in those cases where the standard Wald statistic fails. Although the alternative Wald statistics are derived for tests of Granger noncausality at higher forecast horizons, they may be suitably modified to test for some other null hypothesis with similar nonlinear restrictions. In particular, they may be used with impulse response analysis.

To gain insight into the small sample size and power properties of the alternative Wald statistics proposed in this study, relative to the standard Wald statistic, a small simulation study is set up. The description of the simulation study and the presentation and discussion of the results are given in Chapter 6. A summary and final conclusions follow in Chapter 7.

Chapter 2

Model Setup

In principle, Granger causality and impulse response analysis can be defined without reference to a specific (linear) model.[1] However, when it comes to testing, the restrictions for noncausality have to be explored on the basis of a specific model, and different models lead to different restrictions.

The class of models considered in this chapter are stationary and nonstationary, linear vector autoregressive, discrete time processes.[2]

Studies of Granger causality in continuous time processes can be found in Florens & Fougère (1996) and Comte & Renault (1996). Moreover, causality is defined in time domain. For studies of Granger causality in frequency domain see Geweke (1982), Hosoya (1991) and Granger & Lin (1995). The vector autoregressive models are assumed to be of finite order p with p known a priori. In practice, the true lag order will hardly be known, but can be determined with consistent order selection criteria (see Lütkepohl (1991, Proposition 4.2)). Although the problem of estimating the true lag order arises with any inference in VAR models, it will be neglected in what follows. Studies of Granger causality and impulse response analysis in infinite order VAR models can be found in Lütkepohl (1996b), Lütkepohl & Poskitt (1996) and Lütkepohl & Saikkonen (1997).

---

[1] See for example Hosoya (1977), Florens & Mouchart (1985) and Dufour & Renault (1994, 1998) for definitions of Granger causality in Hilbert spaces.

[2] Results for stationary VAR models can be found e. g. in Lütkepohl (1991) and Hamilton (1994). Nonstationary VAR models are treated for instance in Lütkepohl (1991), Banerjee et al. (1993), Johansen (1995) and Hatanaka (1996).

2.1 Vector Autoregressive Representation

Let y_t be a k–dimensional random vector generated by a vector autoregressive model of finite order p:

$$\Pi(L)\, y_t = \nu + u_t, \quad t = 1, 2, \ldots, \tag{2.1}$$

with

y_t	:	a zero–mean $(k \times 1)$ vector of stochastic variables $y_{1,t}, \ldots, y_{k,t}$,
u_t	:	an independently and identically distributed vector of error terms with $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$ a nonsingular covariance matrix and $E(u_t u_s') = 0$ for $t \neq s$,
ν	:	a $(k \times 1)$ vector of constants,
$\Pi(L)$	:	a matrix polynomial in the $(k \times k)$ fixed coefficient matrices Π_i, $i = 1, \ldots, p$, and the lag operator L, i.e. $\Pi(L) = I_k - \Pi_1 L - \Pi_2 L^2 - \ldots - \Pi_p L^p$ and I_k the $(k \times k)$ identity matrix,
L	:	the lag operator which shifts y_t back in time so that $Ly_t = y_{t-1}$,
$y_{-p+1}, \ldots, y_0$	:	the initial values which are assumed to be fixed, and
p	:	the order of the VAR model.

A variable $y_{i,t}$ of the vector y_t is weakly stationary if it possesses a finite and constant mean and variance and if the autocovariance of $y_{i,t}$ and $y_{i,s}$ for different time periods t, s is finite and depends only on the difference $t - s$. Stationary variables are characterized by fluctuations around their mean. However, economic time series often exhibit trend behaviour. A variable $y_{i,t}$ is called trend stationary if subtracting a deterministic trend function renders it weakly stationary. If $y_{i,t}$ cannot be rendered stationary by substracting deterministic terms but the differenced series $\Delta y_{i,t} = y_{i,t} - y_{i,t-1}$ is stationary, then $y_{i,t}$ is called integrated of order one and denoted $y_{i,t} \sim I(1)$ (see Engle & Granger (1987), Lütkepohl (1991, p. 346), Banerjee et al. (1993, p. 6), Johansen (1995, p. 35)). Variables which are integrated of order one are often said to contain a stochastic trend.

If $y_{i,t} \sim I(1)$, then the differenced series $\Delta y_{i,t}$ is called integrated of order zero and denoted $\Delta y_{i,t} \sim I(0)$ in the following. An $I(0)$ variable is stationary.[3] Hence, an $I(0)$ variable is characterized by weak stationarity and by the fact that "integrating" the series yields an $I(1)$ variable (see Banerjee et al. (1993, Chapter 3)).

The $(k \times 1)$ vector $y_t = [y_{1,t}, \ldots, y_{k,t}]'$ will be called integrated of order zero and denoted $y_t \sim I(0)$ in the following if at least one variable in y_t is of order $I(0)$ and the remaining variables are of the same or of a lower order $I(d)$, $d \in \{0, -1, -2, \ldots\}$. The vector y_t will be called integrated of order one and denoted $y_t \sim I(1)$ if at least one variable in y_t is integrated of order one and the other variables are integrated of the same or of a lower order (see for example Warne (1990), Lütkepohl (1991, p. 346) Johansen (1995, Chapter 3)).

If y_t is $I(1)$ and there exists a $(k \times 1)$ vector b_1 so that the linear combination $b_1' y_t$ is stationary, then b_1 is called a cointegrating vector. To illustrate this definition of cointegration, assume that the vector y_t consists of only two variables $y_{1,t}$ and $y_{2,t}$ where $y_{1,t} \sim I(1)$ and $y_{2,t} \sim I(d)$, $d \in \{1, 0, -1, \ldots\}$. Then cointegration can exist in the following two situations:

(a) $y_{2,t}$ is $I(1)$ and $b_1 = [b_{11}, b_{21}]'$ with $b_{11}, b_{21} \neq 0$, and

(b) $y_{2,t} \sim I(d)$, $d \in \{0, -1, \ldots\}$, and $b_1 = [0, b_{21}]'$ with $b_{21} \neq 0$.

In case (a), $b_1' y_t = b_{11} y_{1,t} + b_{21} y_{2,t}$ is $I(0)$ if there exists a linear combination of integrated variables which is integrated of a lower order. In contrast, in case (b), $b_1' y_t = b_{21} y_{2,t}$ describes only a linear transformation of a stationary variable. Note, that the definitions of integration and cointegration used here are wider than the ones given in Engle & Granger (1987). They are preferred here because they allow the joint analysis of stationary and integrated variables.[4]

---

[3] However, not every stationary series is $I(0)$. For instance, differencing an $I(0)$ variable yields again a stationary variable which may be denoted $I(-1)$.

[4] Engle & Granger (1987) define a k–dimensional vector of variables y_t to be integrated of order one if each element $y_{i,t}$ of y_t is integrated of order one for $i = 1, \ldots, k$. Consequently, their definition of cointegration is limited to case (a).

If the determinantal polynomial $det(\Pi(z)) = det(I_k - \Pi_1 z - \Pi_2 z^2 - \ldots - \Pi_p z^p)$ fulfills the condition

$$det(\Pi(z)) \neq 0 \quad \text{for } |z| \leq 1, \tag{2.2}$$

i.e. if the determinantal polynomial has no roots with modulus less than or equal to one, the vector autoregressive process in (2.3) on page 8 is stable and the matrix polynomial $\Pi(L)$ is invertible. Furthermore, stability implies stationarity of the variables in y_t.[5]

If the stability condition (2.2) is not fulfilled, the vector autoregressive process is nonstationary. In this case, the determinantal polynomial may have explosive roots ($|z| < 1$), seasonal roots ($|z| = 1$ but $z \neq 1$) or unit roots ($z = 1$) (Johansen (1995, p. 14)).

Another source of nonstationarity are deterministic trend functions. Whenever the terms nonstationary and nonstationarity are used in this thesis, they refer to nonstationarity which arises from unit roots of the determinantal polynomial. As consequence, nonstationarity of a vector autoregressive process implies here that at least one variable in y_t is integrated of order one.

For simplicity of exposition, the deterministic term ν will be set to zero in the following as this term plays no role in defining causality respectively noncausality.[6] Hence, the following VAR process

$$\Pi(L)\, y_t = u_t, \quad t = 1, 2, \ldots, \tag{2.3}$$

is considered in subsequent chapters.

2.2 Error Correction Representation

The matrix polynomial $\Pi(L)$ can be factored into

$$\Pi(L) = \Pi(1)L + (1 - L)\Gamma(L), \tag{2.4}$$

[5]Strictly speaking, stability implies only asymptotic stationarity. A common assumption is therefore that the vector autoregressive process started in the infinite past. If in contrast the initial values are assumed to be fixed, i.e. if the process is conditioned on the initial values, stability implies stationarity only under the additional assumption that the vector of initial variables y_0 has the same distribution as the unconditional process (Lütkepohl (1991, Chapters 2 and 11), Hamilton (1994, Chapter 10), Johansen (1995, p. 15)). The latter assumption is assumed to hold here.

[6]Note, however, that deterministic terms play an important role in modeling the data and in estimation. Moreover, critical values of unit root tests and tests for the cointegration rank depend on whether deterministic terms are included in the regression equation or not.

where

$$\begin{aligned}\Gamma(L) &= I_k - \Gamma_1 L - \ldots - \Gamma_{p-1} L^{p-1},\\ \Gamma_j &= -\sum_{i=j+1}^{p} \Pi_i,\\ \Pi(1) &= I_k - \sum_{i=1}^{p} \Pi_i\end{aligned}$$

(see e.g. Johansen (1995)). Inserting (2.4) into (2.3) yields the error correction (EC) representation:

$$\Gamma(L)\,\Delta y_t = -\Pi(1)\,y_{t-1} + u_t, \tag{2.5}$$

where $\Delta y_t = (1-L)y_t$. Rearranging terms yields

$$\Delta y_t = -\Pi(1)\,y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j\,\Delta y_{t-j} + u_t. \tag{2.6}$$

If $y_t = [y_{1,t}, \ldots, y_{k,t}]'$ is integrated of order zero, then the differenced vector $\Delta y_t \sim I(-1)$ is stationary. In this case, only stationary variables enter into equation (2.6) and the matrix $\Pi(1)$ has full rank k.

If y_t contains variables which are integrated of order one and possibly cointegrated, $y_t \sim I(1)$ while differencing the vector once yields a vector of stationary variables $\Delta y_t \sim I(0)$. Since the left–hand side of equation (2.6) contains only variables integrated of order zero, the same must be true for the right–hand side of the equation. This implies that $\Pi(1)\,y_{t-1}$ is either zero or stationary:

If $y_t \sim I(1)$, the determinantal polynomial in (2.2) has at least one unit root $z = 1$. As consequence, the condition $det(\Pi(1)) = 0$ implies that $\Pi(1)$ does not have full rank k but has a reduced rank r with $0 \leq r < k$. If $r = 0$, then $\Pi(1) = 0$, and the error correction representation reduces to a vector autoregressive model in first differences. For $0 < r < k$, there exist $(k \times r)$ matrices A, B of rank r such that

$$\Pi(1) = AB' \tag{2.7}$$

where $B'y_{t-1}$, and hence $\Pi(1)\,y_{t-1}$, are stationary (see Granger (1986), Engle & Granger (1987)). The matrix B contains r column vectors b_s, $s = 1, \ldots, r$, which together with the k variables in y_t form r linearly independent, stationary combinations $b_s'y_t$. These are called the cointegrating relations. A set of variables in y_t is said to be cointegrated,

if there exists a linear combination $b_s' y_t$ which is stationary.

Stationary variables fluctuate around their unconditional mean while variables which are integrated of order one can wander widely without returning to their mean value in finite time. However, if two or more integrated time series are cointegrated, they form a stationary relationship and hence are tied together in the long–run. Based on this interpretation, the matrix $\Pi(1)$ is often called the long–run matrix. On the other hand, the coefficient matrices Γ_j which are attached to the stationary variables in Δy_{t-j} are said to measure the short–run effects. In the literature, special interest has been directed towards the long–run relationship $\Pi(1)\, y_{t-1} = AB' y_{t-1}$.

Note, that the matrices A, B are not unique unless normalized. For any nonsingular matrix F, $\tilde{A} = AF$ and $\tilde{B}' = F^{-1}B'$ represents another admissible set of matrices. It is therefore possible to normalize the matrices in such a way that a cointegrating relation $b_s' y_{t-1}$ can be interpreted as deviation from an equilibrium relation between the variables involved. To illustrate this, assume that the vector y_t contains only two variables $y_{1,t}$, $y_{2,t} \sim I(1)$ which are cointegrated. Since $k = 2$, it follows from $0 < r < k$ that there is only one cointegrating vector b_1. Let this vector be normalized as $b_1' = [1,\, b_{21}]$, then $b_1' y_{t-1} = y_{1,t-1} + b_{21}\, y_{2,t-1}$ can be interpreted as deviation from the long–run equilibrium $y_{1,t} = -b_{21}\, y_{2,t}$, or as equilibrium error.

Since economic theory usually has more to say about equilibrium relations than about short–run dynamics, it may help in the normalization. Stationarity of $B' y_{t-1}$ ensures that the deviations from the r long–run (equilibrium) relations fluctuate around their mean value of zero. The matrix A measures how much of the deviations from the equilibrium relations is corrected for in the next period; hence the name error correction representation. Note, however, that this interpretation rests on the definition of cointegration as a long–run relationship between two or more integrated variables.

In this thesis, a vector y_t is $I(1)$ even if some of its elements are stationary. Consequently, a cointegrating relation $b_s' y_t$ may describe a long–run relationship between two or more integrated variables, a linear combination of stationary variables or even just one stationary variable.

2.3 Moving Average Representation

If the stability condition (2.2) on page 8 is satisfied, the polynomial matrix $\Pi(L)$ can be inverted, i.e. there exists a polynomial matrix $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$, such that

$$\Pi(L)\,\Phi(L) = I_k. \tag{2.8}$$

Inversion of the stable, stationary VAR(p) process defined by (2.3) and (2.2) yields the moving average (MA) representation:

$$y_t = \Phi(L)\,u_t, \quad t = 1, 2, \ldots \tag{2.9}$$

The coefficient matrices Φ_j of the moving average polynomial $\Phi(L)$ can be obtained recursively from the vector autoregressive coefficient matrices as follows:

$$\Phi_j = \sum_{i=1}^{j} \Phi_{j-i}\,\Pi_i \quad \text{and} \quad \Phi_0 = I_k. \tag{2.10}$$

The vector of error terms u_t represents the error of the optimal one–step ahead forecast of y_t at time $t-1$. Moreover, representation (2.9) expresses y_t as a linear function of these forecast errors. It is therefore sometimes called the prediction error representation (see e.g. Lütkepohl (1991, Chapter 2)).

If y_t contains some variables which are integrated of order one, the vector autoregressive polynomial matrix $\Pi(L)$ is no longer invertible. Although one may still compute coefficient matrices Φ_j according to the recursion formula (2.10), these matrices may not converge to zero for $j \to \infty$ (see Lütkepohl (1991, p. 380), Lütkepohl & Breitung (1997, p. 303)). They can therefore not be called moving average coefficient matrices as y_t does no longer possess a convergent moving average representation.
If $y_t \sim I(1)$, then $\Delta y_t \sim I(0)$. It then follows from Wold's decomposition theorem that the zero–mean stationary process Δy_t has an infinite order moving average representation (see Lütkepohl (1991, Chapter 2), Hamilton (1994, Chapter 4)):

$$\Delta y_t = C(L)\,u_t, \quad t = 1, 2, \ldots, \tag{2.11}$$

with

$$C(L) = \sum_{s=0}^{\infty} C_s L^s \quad \text{and} \quad C_0 = I_k. \tag{2.12}$$

For finite t and initial values $y_0 = y_{-1} = y_{-2} = \ldots = 0$, the level variables y_t can be computed as

$$y_t = \sum_{j=0}^{t-1} \Delta y_{t-j} = \sum_{j=0}^{t-1} C(L)\, u_{t-j} = \sum_{j=0}^{t-1} \sum_{s=0}^{j} C_s\, u_{t-j} = \sum_{j=0}^{t-1} \Phi_j\, u_{t-j}. \tag{2.13}$$

Hence, alternatively to the recursion formula (2.10), the coefficient matrices Φ_j can be computed as

$$\Phi_j = \sum_{s=0}^{j} C_s. \tag{2.14}$$

The latter formula illustrates why the coefficient matrices Φ_j may not converge for $j \to \infty$ if $y_t \sim I(1)$.

Although the recursion formula (2.10) does not depend on the order of integration of the variables in y_t, the resulting coefficient matrices Φ_j have different properties in stationary than in nonstationary vector autoregressive models. Chapter 3 addresses this point in the context of impulse response analysis.

2.4 Common Trends Representation

Starting from the moving average representation in (2.11), the polynomial $C(L)$ can be rewritten as

$$C(L) = C(1) + (1 - L)C^{\star}(L) \tag{2.15}$$

with

$$\begin{aligned} C(1) &= I_k + \sum_{s=1}^{\infty} C_s, \\ C^{\star}(L) &= \sum_{j=0}^{\infty} C_j^{\star} L^j, \\ C_j^{\star} &= -\sum_{s=j+1}^{\infty} C_s, \\ C_0^{\star} &= I_k - C(1). \end{aligned}$$

The existence of these matrices is guaranteed under the assumption that the C_s matrices in (2.12) obey an exponential decay condition so that $\sum_{s=0}^{p} C_s L^s$ is a convergent

sequence as $p \to \infty$ (see Banerjee et al. (1993, p. 257) and Johansen (1995, p. 14)).

Let $y_t \sim I(1)$ and cointegrated with cointegration rank $0 < r < k$ and let $\Pi(1) = AB'$, then it follows from Granger's Representation Theorem, that y_t has a common trends (CT) representation

$$y_t = C(1) \sum_{s=1}^{t} u_s + C^\star(L)\, u_t + y_0 - C^\star(L)\, u_0, \tag{2.16}$$

where

$$\begin{aligned} C(1) &= B_\perp (A_\perp' \Gamma(1) B_\perp)^{-1} A_\perp', \\ \Gamma(1) &= I_k - \sum_{j=1}^{p-1} \Gamma_j, \end{aligned}$$

and $A_\perp, B_\perp$ are $(k \times (k-r))$ matrices of full column rank which fulfill the conditions $A_\perp' A = 0$ and $B_\perp' B = 0$. Since the matrices $A_\perp, B_\perp$ have column rank $(k-r)$, it follows that the $(k \times k)$ matrix $C(1)$ has reduced rank equal to $(k-r)$ with $0 \leq r < k$ (see Granger (1986), Engle & Granger (1987), Johansen (1995, Theorem 4.2)).

Similar to the error correction representation, the common trends representation allows to distinguish between long–run and short–run terms: in particular, y_t is decomposed into a long–run component $C(1) \sum_{s=1}^{t} u_s$, a short–run component $C^\star(L)\, u_t$ and the initial conditions $y_0 - C^\star(L)\, u_0$.

One is often interested in the long–run component

$$C(1) \sum_{s=1}^{t} u_s \;=\; B_\perp (A_\perp' \Gamma(1) B_\perp)^{-1} A_\perp' \sum_{s=1}^{t} u_s.$$

The accumulation of the stationary error terms in $u_t = [u_{1,t}, \ldots, u_{k,t}]'$ over t periods generates stochastic trends. The matrix $C(1)$ contains the information about how the stochastic trends are weighted to $(k-r)$ linear combinations $A_\perp' \sum_{s=1}^{t} u_s$, the so–called common trends. The matrix $B_\perp (A_\perp' \Gamma(1) B_\perp)^{-1}$ shows how these common trends enter the system.

If y_t contains only stationary variables, $C(1) = 0$ and the common trends representation reduces to the moving average representation (2.9), see page 11.

2.5 Companion Form

Using the companion form, it is possible to write a vector autoregressive model of arbitrary finite order p as a VAR(1) model

$$Y_t = \mathbf{\Pi}\, Y_{t-1} + U_t, \quad t = 1, 2, \ldots, \tag{2.17}$$

with

$$Y_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+1} \end{bmatrix} \qquad (kp \times 1),$$

$$\mathbf{\Pi} = \begin{bmatrix} \Pi_1 & \ldots & \Pi_{p-1} & \Pi_p \\ I_k & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & I_k & 0 \end{bmatrix} \qquad (kp \times kp),$$

$$U_t = \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (kp \times 1).$$

Let $J = [I_k, 0, \ldots, 0]$ be a $(k \times kp)$ matrix with I_k the identity matrix, then left–multiplying Y_t with J yields the VAR representation (2.3) on page 8 (see Lütkepohl (1991, Chapter 2), Hamilton (1994, Chapter 10)).

Chapter 3

Causality in VAR Models: Representation.

Granger causality is a concept used to study causality between time series variables. Let $y_{1,t}$, $y_{2,t}$ be realizations of two stochastic time series y_1, y_2 at time point t, then Granger defines y_1 to be causal for y_2 if the information in the past and present of y_1 at time t helps to improve the one–step ahead forecast of $y_{2,t}$. If the information in $y_{1,t}$, $y_{1,t-1}$, ... cannot help to predict $y_{2,t+1}$, y_1 is called Granger noncausal for y_2. This definition of causality rests on the two main assumptions that the cause precedes the effect in time and that the causal series y_1 contains some unique information about y_2 (see Granger (1988)).

The assumption that the cause precedes the effect in time seems plausible at first sight. However, it is not sufficient to identify a cause in a multiple time series context. Indeed, there may exist a lot of different time series which can help forecast another time series although there is no logical cause–effect relationship between them. For instance Sheehan & Grieves (1982) find a Granger causal relationship between sunspots and business cycles. Granger causality occurs because both time series show a similar cyclical pattern and hence correlate. The Sheehan & Grieves study underlines that correlation plus improved predictability is not sufficient to identify a cause–effect relationship: the time series which enter a Granger causality study should justify a potential causal relationship on some theoretical grounds. At least, commonsense should rule out some

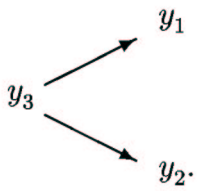
causal relationships.

If all information available in the universe had been used in the Sheehan & Grieves study, it is presumably likely that other time series would explain business cycles better than sunspots.[1] Put differently, sunspots do not have unique information about business cycles and hence the condition that a cause should have some unique information which cannot be found in other time series is violated. However, in practice a Granger causal study cannot take into account all information available in the universe but will rather be limited to a small number of variables. This stresses once again the importance to carefully define the set of relevant variables.

A different problem arises from the fact that some causes are unobservable: for instance, we are not able to observe the expectations of economic agents which cause economic actions, but only the actions themselves. Now using the criterion that the cause must precede the effect may lead to a reversed causal direction (see e.g. Gupta (1987)). A little example may illustrate this problem: if a price increase of a good is anticipated, consumers might increase their demand for the good (y_2) to build stocks before the price (y_1) will rise. We therefore observe a causal link running from increased demand (y_2) to price increase (y_1) although in fact y_2 is caused by expected future y_1. A similar mechanism works in the case of control variables and leading indicators. This type of *spurious causality* may also be due to measurement errors if these errors have a certain time structure. Examples are given in Sims (1972), Granger (1980), Hsiao (1982), Newbold (1982), Hamilton (1994, Chapter 11), and Leamer (1985).

Granger causal links are sensitive to the information set which is employed in the analysis. Changing the information set, for example by extending or reducing the number of time series in the study, may lead to different Granger causal links. One example are common causes: let y_3 be a third variable containing information which helps to predict y_1 as well as y_2, illustrated as

[1] If, however, economic agents believe that sunspots cause indeed business cycles, their anticipations work like a self–fulfilling prophecy, turning sunspots into a true cause. This case is excluded here.



In this case it is possible to find Granger causality of y_1 for y_2 in a bivariate VAR model although y_1 is not Granger causal for y_2 if the information in y_3 is taken into account. Spurious causality arises in the bivariate VAR model because y_1 picks up part of the left–out information in y_3. Hence spurious causal links arise whenever relevant information is left out but is picked up by the included variables (see Schneider (1991), Hamilton (1994, Chapter 11)).

In vector autoregressive models this problem gains importance in small samples: since the number of coefficients grows with the square of the number of variables, there is a trade–off between inclusion of another (relevant) variable and estimation uncertainty (Geweke (1984, p. 1140f)). For further work on how omitted variables affect Granger causal links see e.g. Lütkepohl (1982), Florens & Mouchart (1982, 1985), Braun & Mittnik (1993), Caporale & Pittis (1997), Caporale et al. (1998) and Triacca (1998).

The fact that Granger causal relationships depend on the information set which is taken into account in the analysis has stimulated a thorough discussion on how Granger causality is related to philosophical definitions of causality. The latter definitions agree on causality as an *invariant and necessary relationship between cause and effect.* A causal relationship should therefore be unaffected by changes of the information set. In particular, the definition of Feigl (1953) who defines causality as *predictability according to a law or a set of laws* stresses that causality is understood as a regularity of the real world, following a *law* in the sense of physical laws rather than statistical regularities. A survey on this discussion can be found e.g. in Zellner (1979), Geweke (1984), Conway et al. (1984), Cooley & Leroy (1985), Holland (1986), Basmann (1988), Cox (1992), Hillmer (1992) and Vercelli (1992).

While physical laws hold always with certainty, Granger causality characterizes a sta-

tistical relationship between time series in a specific time period which may not hold in other time periods unless *all causal relationships remain constant in direction through time* (Granger (1980, Axiom C)). The latter assumption is violated in practice if the data generating mechanism changes over time. Structural breaks are a well–known example (Lütkepohl (1989a)). Zellner (1979) therefore speaks of the *post hoc ergo propter hoc* fallacy, see also Tobin (1970). As consequence, a Granger causal relationship may not hold in time periods which have not been considered in the analysis.

Physical laws allow for instantaneous causality. However, a Granger causal structure is identified by the time lag between the causal variable y_1 and the effect y_2, and therefore the concept of Granger causality excludes instantaneous causality: if the information in $y_{1,t+1}$ helps to improve the one–step ahead prediction of $y_{2,t}$, there is contemporaneous correlation between y_1 and y_2 but no time lag to identify the causal direction. In this case, an instantaneous causal direction can only be derived from prior information. While Granger (1969, 1980, 1988) and Granger & Newbold (1986) doubt that there exists true instantaneous causality, at least among economic time series, they admit that instantaneous causality may spuriously arise in a model due to temporal aggregation or common causes. This distinguishes the concept from other econometric concepts where assumptions on true instantaneous causal links play a role: for example, assumptions about instantaneous causality are used to model the contemporaneous relations among the endogenous variables of simultaneous equations models. Moreover, they play a role in orthogonal impulse response analysis, forecast error decomposition and impulse response analysis in structural VAR models (Lütkepohl (1991), Breitung (1998), Lütkepohl (1999)).

In economics, it seems hard to agree on immutable laws, but much easier to agree on statistical regularities or stylized facts. Granger causality tests can be used to detect relationships between economic time series and to create stylized facts. Granger causality tests have become popular for several reasons: they are easy to implement in general; they are based on a simple but appealing concept which gains transparency by eschewing prior information; the vivid discussion on the concept of Granger causality

as well as on Granger causality tests has already revealed its merits and pitfalls; the ongoing discussion continuously stimulates research on more sophisticated versions of Granger causality and testing procedures.

One of these research areas is the non–robustness of Granger causal links with respect to the information set used in the causality analysis. Thereby, special attention has been directed to the fact that data may be observed or reported at time points different from their generation date:
Data are often published as averages, sums or end–of–period data. If data are temporally aggregated, they are available at a uency than they are generated. Since temporal aggregation changes the time structure of the variables, it affects their information content. For instance, forecasting an aggregated time series with aggregated data differs in general from forecasting with disaggregated data and aggregating the forecast (Lütkepohl (1986) and Lütkepohl (1989b)). Similarly, Granger causal links may depend on whether disaggregated or aggregated data are used:
Assume for instance that $y_{1,t}$, $y_{2,t}$ are two price series generated at monthly frequency and that forecasts of $y_{2,t+1}$ can be improved if the information in $y_{1,t}$ is used. In other words, y_1 Granger causes y_2 at a monthly frequency. Now if only quarterly data are reported, a Granger causal link between $x_{1,s} = y_{1,t} + y_{1,t+1} + y_{1,t+2}$ and $x_{2,s+1} = y_{2,t+3} + y_{2,t+4} + y_{2,t+5}$ may not be detected: although y_1 Granger causes y_2, this relationship is weakened by the temporal aggregation. Instead, instantaneous causality between x_1 and x_2 arises because $x_{1,s}$ ($x_{2,s}$) contains $y_{1,t}$ ($y_{2,t+1}$), which are correlated (Granger (1969, 1980)).
The simple example already illustrates that a causal relationship between two original time series need not hold between the transformed time series, even though identical, linear transformations are used (see e.g. Pierce & Haugh (1977), Kirchgässner (1981), Tiede (1991), Kirchgässner & Wolters (1992)). This result gains relevance in empirical applications where data are commonly transformed in various ways to adjust for seasonality, outliers and deterministic or stochastic trends.

Several authors have tried to derive conditions under which Granger causality in the

model with aggregated data also implies Granger causality in the disaggregated model (see e.g. Comte & Renault (1996), Florens & Fougère (1996), Dufour & Renault (1998), Renault et al. (1998), Breitung & Swanson (1998)). Robustness of Granger causality with respect to (dis)aggregation is important if data are assumed to be generated at a higher frequency than they are reported:

For instance, Comte & Renault (1996) and Renault et al. (1998) analyze the robustness of Granger causal links in a discrete time model when the true underlying model is either a continuous time model or a discrete time model observed at a higher frequency. In both cases the observation period is fixed but the data frequency and hence the number of observations increases. In contrast, Breitung & Swanson (1998) analyze robustness of Granger causality in a discrete time model where the number of observations is held constant but the aggregation interval goes to infinity.

The different contributions show that temporal robustness of Granger causality holds for special cases but not in general.

Further work has concentrated on the extension of the original concept of Granger (1969) to higher forecast horizons $h > 1$. This extension is useful if relationships among three or more time series are studied:

If the analysis of Granger causality is limited to only two time series y_1, y_2, Granger causality of y_1 for y_2 must show up at forecast horizon one: either y_1 has some unique information about y_2 or it has not. If y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$, it will also be Granger noncausal for y_2 at higher forecast horizons $h > 1$ (Pierce (1975)).

However, if there is a third variable y_3, y_1 may be Granger noncausal for y_2 at forecast horizon $h = 1$ but indirectly cause y_2 at higher forecast horizons $h > 1$ (see Granger (1980), Newbold (1982), Lütkepohl (1982)). This *indirect causality* occurs if y_1 has some unique information which helps to predict y_3 at forecast horizon $h = 1$, while y_3 Granger causes y_2 in the following period. Indirect causality of y_1 for y_2 then shows up at forecast horizon $h = 2$:

$$y_1 \xrightarrow{1} y_3 \xrightarrow{1} y_2$$

In general, indirect causality of y_1 for y_2 may show up at some higher forecast horizon

$h \geq 2$. The actual length of an indirect causal chain depends on the number of third variables and the lag structure of the underlying vector autoregressive model.

Different names have been suggested to distinguish Granger causality at forecast horizon $h > 1$ from standard Granger causality at forecast horizon $h = 1$. Examples are *long–run* or *multiple–horizon* causality by Dufour & Renault (1994, 1998), *weak global causality* by Florens & Fougère (1996) and *multi–step causality* by Lütkepohl & Burda (1997). In this thesis, the extension of the standard Granger causality concept to higher forecast horizons $h \geq 1$ is called extended Granger causality. This concept encompasses standard (direct) Granger causality ($h = 1$) and indirect Granger causality ($h > 1$). To ensure that a variable y_1 neither directly nor indirectly Granger causes another variable y_2, Granger causality has to be checked at all forecast horizons $h \geq 1$. However, in finite order VAR(p) models, it suffices to check only a finite number of forecast horizons to ensure that y_1 is never Granger causal for y_2. The restrictions for Granger noncausality at all forecast horizons $h \geq 1$ have been explored e.g. by Boudjellaba et al. (1992a, 1992b), Dufour & Tessier (1992), Bruneau & Nicolai (1992a), Bruneau & Nicolai (1994) Lütkepohl (1993) and Dufour & Renault (1994, 1998).

Impulse response analysis offers an alternative way to analyze causal relationships among time series in vector autoregressive models. Thereby, a causal relationship is interpreted as a stimulus–response–mechanism: if y_1 is causal for y_2, a shock in y_1 in time period t should stimulate a response of y_2 in future time periods $t+1$, $t+2$, ... The impulse is often modeled as a one–time one–unit exogenous shock which only affects y_1 at time t. If this new information does not change the forecasts of $y_{2,t+h}$ for all forecast horizons $h \geq 1$, then y_1 may be called noncausal for y_2. If the forecast of $y_{2,t+h}$ responds to an impulse in $y_{1,t}$, one might want to gain an idea about the size and duration of the responses. Therefore the responses can be summed up over h periods to give the h–th interim multiplier. Moreover, the long–run effect can be computed by accumulating all responses, at least in stationary VAR models (see Lütkepohl (1991, Chapter 2)).

Standard Granger causality and impulse response analysis both rest on the idea that

the cause precedes the effect in time. Moreover, both concepts use the criterion of improved predictability to identify the causal variable (e.g. Sims (1980) and Lütkepohl (1991, Chapter 2)).

Indeed, both concepts may lead to the same set of restrictions: for instance in vector autoregressive models with only two time series y_1, y_2, standard Granger noncausality of y_1 for y_2 implies and is implied by zero responses of y_2 to an impulse in y_1. On the other hand, standard Granger causality implies that some impulse responses are nonzero (see Sims (1972), Hosoya (1977), Feige & Pearce (1979)). In the bivariate VAR model, the one–time impulse in $y_{1,t}$ changes the value of $y_{2,t+h}$ in subsequent periods $h \geq 1$ only, if the information in $y_{1,t}$, $y_{1,t-1}$, ... helps to better predict $y_{2,t+h}$, hence if y_1 Granger causes y_2.

In VAR models with a vector of third variables (y_3), y_1 may be Granger noncausal for y_2 at forecast horizon $h = 1$, and yet an exogenous shock in $y_{1,t}$ may lead to nonzero responses of $y_{2,t+h}$ at higher forecast horizons $h > 1$ (see e.g. Lütkepohl (1982, 1993), Dufour & Tessier (1992)). This situation arises if an impulse in y_1 leads to a nonzero response of some third variable in y_3 which then stimulates a response of y_2 in later periods. Hence, like the extended Granger causality concept, impulse response analysis takes into account indirect causal chains while the standard Granger causality concept considers only direct causal links.

The subsequent chapter is divided into two parts: in the first part, the concepts of standard Granger causality, extended Granger causality and impulse response analysis are presented and noncausality restrictions under either concept are derived within the stationary VAR(p) model framework. Thereby, the main focus rests on illustrating the concept of extended Granger causality relative to standard Granger causality and impulse response analysis. In the second part, it is shown that the restrictions for noncausality carry over to nonstationary VAR models. Moreover, related causality concepts like long–run and short–run Granger causality or neutrality are presented to complete the overview.

3.1 Causality in Stable, Stationary VAR Models

3.1.1 Standard and Extended Granger Causality

Since stationary and nonstationary variables have quite different properties, restrictions for Granger noncausality at forecast horizons $h = 1$ and $h > 1$ and for noncausality in terms of impulse response analysis are derived first for the stable, stationary VAR model in which the $(k \times 1)$ vector y_t consists of stationary variables only. It is assumed that the k–dimensional vector y_t can be partitioned into $y_{1,t}$, $y_{2,t}$, $y_{3,t}$ of dimensions k_1, k_2, k_3 with $k = k_1 + k_2 + k_3$, where k_1, $k_2 \geq 1$ and $k_3 \geq 0$. A corresponding partition of the vector autoregressive coefficient matrices yields the following representation:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \pi_{11,i} & \pi_{12,i} & \pi_{13,i} \\ \pi_{21,i} & \pi_{22,i} & \pi_{23,i} \\ \pi_{31,i} & \pi_{32,i} & \pi_{33,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \\ y_{3,t-i} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}. \tag{3.1}$$

Thereby, $\pi_{fg,i}$ denotes a $(k_f \times k_g)$ coefficient matrix.

Throughout this thesis, interest centers on whether y_1 is causal for y_2. Although some causality concepts allow the study of causality between vectors (k_1, $k_2 > 1$), interest usually focuses on causality between variables. Therefore, $k_1 = k_2 = 1$ in the following if not mentioned otherwise.

VAR models often contain more than only two variables. These $k_3 = k - (k_1 + k_2)$ third variables are contained in the vector y_3. If $k_3 = 1$, y_3 reduces to one variable. If $k_3 > 1$, y_3 is a vector of k_3 third variables, denoted as $y_3 = [\mathtt{y}_{\mathtt{k}_1+\mathtt{k}_2+\mathtt{1}}, \ldots, \mathtt{y}_{\mathtt{k}}]'$. For $k_3 = 0$ there are no third variables, and the partition reduced to a bivariate VAR model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \sum_{i=1}^{p} \begin{bmatrix} \pi_{11,i} & \pi_{12,i} \\ \pi_{21,i} & \pi_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-i} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}. \tag{3.2}$$

Assume that the k–dimensional vector $y_t = [y_{1,t}, y_{2,t}, y'_{3,t}]'$ is generated by the stable, stationary VAR model (3.1). If y_1 is (directly) Granger causal for y_2, the information in the past and present of $y_{1,t}$ should help to improve the one–step ahead forecast of $y_{2,t}$. To measure improved predictability, a loss function has to be specified first. A common loss function is the forecast mean–squared error (MSE). The point forecast which minimizes this loss function is the expectation of $y_{2,t+1}$, conditional on the information set I_t

which has been used in the forecast. The conditional expectation requires knowledge of the conditional density function which may not be available in practice. Furthermore, the conditional expectation may be a nonlinear function of the information in I_t. To avoid these problems, the forecast function is often restricted to be a linear function of the data in I_t. This restriction leads to a suboptimal predictor in general. However, under the assumption that the error vector u_t in (3.1) is independently identically or even Gaussian distributed, the linear minimum MSE predictor equals the conditional expectation and is hence optimal (see Granger & Newbold (1986)).

A Formal Definition of Standard Granger Noncausality

Let $y_{2,t}(1)$ denote the optimal one–step ahead predictor of y_2 at forecast origin t, which minimizes the MSE among all those predictors which are linear functions of the information in y_t, y_{t-1}, ... Let $y^{\star}_{2,t}(1)$ denote the optimal one–step ahead predictor of y_2 at forecast origin t, which minimizes the MSE among all those predictors which are linear functions of the information in y_t, y_{t-1}, ..., but without the information in $y_{1,t}$, $y_{1,t-1}$, ... The corresponding one–step ahead mean–squared errors are denoted as $\mathrm{MSE}(y_{2,t}(1))$ and $\mathrm{MSE}(y^{\star}_{2,t}(1))$. Then standard Granger causality can be defined as follows.

Definition 3.1: Granger Noncausality at Forecast Horizon $h = 1$.
y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$ iff

$$\mathrm{MSE}(y_{2,t}(1)) \quad = \quad \mathrm{MSE}(y^{\star}_{2,t}(1)).$$

If $k_2 = 1$ and

$$\mathrm{MSE}(y_{2,t}(1)) \quad < \quad \mathrm{MSE}(y^{\star}_{2,t}(1)),$$

then y_1 is said to Granger cause y_2 at forecast horizon 1.
If $k_2 > 1$, Granger causality of y_1 for y_2 requires that the matrices $\mathrm{MSE}(y_{2,t}(1))$, $\mathrm{MSE}(y^{\star}_{2,t}(1))$ are not identical and that the difference $\mathrm{MSE}(y^{\star}_{2,t}(1)) - \mathrm{MSE}(y_{2,t}(1))$ is positive semidefinite (Lütkepohl (1991, Chapter 2)).

Standard Granger Noncausality in the VAR Model

In the vector autoregressive model, equality of the mean–squared errors implies that the optimal linear predictors are also equal. If the optimal linear predictor of $y_{2,t+1}$ does not use the information in the past and present of $y_{1,t}$, then y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$. Let

$$y_t = \sum_{i=1}^{p} \Pi_i \, y_{t-i} + u_t, \quad u_t \sim i.i.d.(0, \Sigma_u),$$

be partitioned as in (3.1) on page 23. The optimal predictor

$$y_{2,t}(1) = \sum_{i=1}^{p} (\pi_{21,i} \, y_{1,t+1-i} + \pi_{22,i} \, y_{2,t+1-i} + \pi_{23,i} \, y_{3,t+1-i})$$

does not use the information in $y_{1,t+1-i}$, $i \geq 1$, if $\pi_{21,i} = 0$ for all $i = 1, \ldots, p$. Granger noncausality is thus characterized by exclusion restrictions on the regressors $y_{1,t-i}$ in the $y_{2,t}$–equations. If on the other hand $\pi_{21,i} \neq 0$ for at least one $i = 1, \ldots, p$, then the predictor $y_{2,t}(1)$ uses more information than the predictor $y^{\star}_{2,t}(1)$ and should therefore yield a smaller mean–squared error (matrix). In this case, y_1 is said to be Granger causal for y_2. This leads to the following proposition:

Proposition 3.1: (Lütkepohl (1991, Proposition 2.2)).
y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$, iff

$$\pi_{21,i} = 0 \quad \text{for } i = 1, \ldots, p.$$

y_1 is said to Granger cause y_2 at forecast horizon $h = 1$, iff

$$\pi_{21,i} \neq 0 \quad \text{for at least one } i \in \{1, \ldots, p\}.$$

Granger causality of y_1 for y_2 at forecast horizon $h = 1$ will be denoted as $y_1 \xrightarrow{1} y_2$ in the following and Granger noncausality as $y_1 \not\xrightarrow{1} y_2$ (see also Boudjellaba et al. (1992b), Dufour & Renault (1994, 1998) and Giles (2000)).

If $y_{1,t}$, $y_{2,t}$ are vectors of variables, Proposition 3.1 states that y_1 is noncausal for y_2 if none of the k_1 components in y_1 is causal for any of the k_2 components in y_2. On the other hand, y_1 is Granger causal for y_2 if at least one element of y_1 Granger causes one element of y_2. Hence, if causality between vectors is found, this does not imply that

all elements of y_1 are causal for all elements of y_2.
If y_1 Granger causes y_2 and y_2 Granger causes y_1, i.e. if $\pi_{21,i} \neq 0$ and also $\pi_{12,i} \neq 0$ holds for at least one $i \in \{1, \ldots, p\}$, then Granger (1969) speaks of a feedback system.

Indirect Causal Chains

Note, that Granger noncausality of y_1 for y_2 does not impose any restrictions on the coefficients in $\pi_{23,i}$ in the $y_{2,t}$–equation(s). However, this does not mean that the information in $y_{3,t-i}$, $i \geq 1$, is irrelevant. Indeed, y_1 may be Granger noncausal for y_2 in a trivariate VAR model but may appear Granger causal for y_2 in a bivariate VAR model where the variables in y_3 are excluded as explanatory variable. This situation arises for example in a trivariate VAR model ($k_3 = 1$), if the left–out variable y_3 Granger causes y_2 and if $y_{1,t-i}$ correlates with $y_{3,t-i}$ for some $i \geq 1$ (see Hamilton (1994, Chapter 11)).

The inclusion of a third variable y_3 offers the possibility that y_1 Granger causes y_3 and y_3 Granger causes y_2:

$$y_1 \xrightarrow{1} y_3 \xrightarrow{1} y_2.$$

This causal chain will not be detected at forecast horizon 1 but at higher forecast horizons. It is therefore a straightforward extension of the standard Granger causality concept to also investigate causality at forecast horizons $h > 1$. The following example illustrates this extended Granger causality concept for the simplest case of a trivariate VAR(1) model:

Example 3.1: Indirect Causal Chains in a Trivariate VAR(1) Model.
Assume that y_t is generated by a VAR(1) model with $k_1 = k_2 = k_3 = 1$, i.e.

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} \pi_{11,1} & \pi_{12,1} & \pi_{13,1} \\ \pi_{21,1} & \pi_{22,1} & \pi_{23,1} \\ \pi_{31,1} & \pi_{32,1} & \pi_{33,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \tag{3.3}$$

respectively

$$y_t = \Pi_1 \, y_{t-1} + u_t, \quad \text{with} \quad u_t \sim \text{i. i. d.}(0, \Sigma_u). \tag{3.4}$$

For a VAR(1) model, the h–step ahead forecast $y_t(h)$ is easily computed as

$$y_t(h) \;=\; \Pi_1^h\, y_t. \tag{3.5}$$

Let Π_1^h be partitioned in correspondence with y_t as

$$\Pi_1^h \;=\; \begin{bmatrix} \pi_{11,1}^{(h)} & \pi_{12,1}^{(h)} & \pi_{13,1}^{(h)} \\ \pi_{21,1}^{(h)} & \pi_{22,1}^{(h)} & \pi_{23,1}^{(h)} \\ \pi_{31,1}^{(h)} & \pi_{32,1}^{(h)} & \pi_{33,1}^{(h)} \end{bmatrix},$$

then the h–step ahead forecast of $y_{2,t}$ can be written as

$$y_{2,t}(h) \;=\; \pi_{21,1}^{(h)}\, y_{1,t} + \pi_{22,1}^{(h)}\, y_{2,t} + \pi_{23,1}^{(h)}\, y_{3,t}.$$

The coefficient $\pi_{21,1}^{(h)}$ reflects the marginal influence of $y_{1,t}$ on $y_{2,t}$ at forecast horizon h (see Lütkepohl (1991, p. 32)). If $\pi_{21,1}^{(h)} \neq 0$, the information in past and present $y_{1,t}$ helps to improve the h–step ahead forecast of $y_{2,t}$ and y_1 is called Granger causal for y_2 at forecast horizon h: $y_1 \xrightarrow{h} y_2$. In contrast, if $\pi_{21,1}^{(h)} = 0$, y_1 is called Granger noncausal for y_2 at forecast horizon h, denoted as $y_1 \overset{h}{\not\longrightarrow} y_2$. If the latter condition holds at all forecast horizons $h = 1, \ldots, \tilde{h}$, then y_1 is said to be Granger noncausal for y_2 *up to* forecast horizon $\tilde{h}$. This is denotes as $y_1 \overset{(\tilde{h})}{\not\longrightarrow} y_2$.

The restrictions for Granger noncausality at higher forecast horizons are nonlinear functions of the vector autoregressive coefficients. For the present example, we obtain the following function for $\pi_{21,1}^{(h)}$ at forecast horizon

$$\begin{aligned}
h = 1: \quad \pi_{21,1}^{(1)} &= \pi_{21,1}, \\
h = 2: \quad \pi_{21,1}^{(2)} &= \pi_{21,1}\pi_{11,1} + \pi_{22,1}\pi_{21,1} + \pi_{23,1}\pi_{31,1}, \\
h = 3: \quad \pi_{21,1}^{(3)} &= \pi_{21,1}\pi_{11,1}\pi_{11,1} + \pi_{21,1}\pi_{12,1}\pi_{21,1} + \pi_{21,1}\pi_{13,1}\pi_{31,1} \\
&+ \pi_{22,1}\pi_{21,1}\pi_{11,1} + \pi_{22,1}\pi_{22,1}\pi_{21,1} + \pi_{22,1}\pi_{23,1}\pi_{31,1} \\
&+ \pi_{23,1}\pi_{31,1}\pi_{11,1} + \pi_{23,1}\pi_{32,1}\pi_{21,1} + \pi_{23,1}\pi_{33,1}\pi_{31,1}.
\end{aligned}$$

If $\pi_{21,1} = 0$, then y_1 does not Granger cause y_2 at forecast horizon $h = 1$. Inserting this restriction yields at forecast horizon

$$\begin{aligned}
h = 2: \quad \pi_{21,1}^{(2)} &= \pi_{23,1}\pi_{31,1}, \\
h = 3: \quad \pi_{21,1}^{(3)} &= \pi_{22,1}\pi_{23,1}\pi_{31,1} + \pi_{23,1}\pi_{31,1}\pi_{11,1} + \pi_{23,1}\pi_{33,1}\pi_{31,1}.
\end{aligned}$$

Hence, at forecast horizon $h = 2$, y_1 has an indirect causal effect on y_2 via y_3, captured in the coefficient product $\pi_{23,1}\pi_{31,1}$. The indirect causal chain is interrupted if either (i) y_1 does not Granger cause y_3 ($\pi_{31,1} = 0$) or (ii) y_3 does not Granger cause y_2 ($\pi_{23,1} = 0$) or (iii) neither y_1 Granger causes y_3 nor y_3 Granger causes y_2 ($\pi_{31,1} = \pi_{23,1} = 0$):

$$\begin{aligned}
&\text{(i) } y_1 \overset{1}{\not\longrightarrow} y_3 \overset{1}{\longrightarrow} y_2,\\
&\text{(ii) } y_1 \overset{1}{\longrightarrow} y_3 \overset{1}{\not\longrightarrow} y_2,\\
&\text{(iii) } y_1 \overset{1}{\not\longrightarrow} y_3 \overset{1}{\not\longrightarrow} y_2.
\end{aligned}$$

Example 3.1 illustrates that in VAR models with third variables, the concept of standard Granger causality does not capture all possible causal links: only direct causal links are taken into account. These direct causal links show up at forecast horizon $h = 1$. But in the presence of third variables, indirect causal chains are possible which run from y_1 to y_3 to y_2 and show up at higher forecast horizons $h > 1$.

Internalizing Indirect Causal Chains: A Separation Criterion

One way to incorporate these indirect causal chains into the standard Granger causality concept is to split up all k variables in y_t into two vectors of variables $S_{1,t}$, $S_{2,t}$ of dimensions s_1 respectively s_2 with $s_1,\ s_2 > 0$ and $k = s_1 + s_2$. Instead of studying Granger causality between all variables in y_t, the analysis is now restricted to Granger causality between the two vectors S_1 and S_2. Since there is no third vector of variables, indirect causal links are not possible. As consequence, Granger noncausality of the vector S_1 for the vector S_2 at forecast horizon $h = 1$ implies Granger noncausality of S_1 for S_2 at all forecast horizons: $S_1 \overset{1}{\not\longrightarrow} S_2 \quad \Longrightarrow \quad S_1 \overset{h}{\not\longrightarrow} S_2 \quad \forall\, h \geq 1$.
Moreover, Granger noncausality of the vector S_1 for the vector S_2 at forecast horizon $h = 1$ implies that all variables in S_1 are neither directly nor indirectly Granger causal for the variables in S_2 (see Dufour & Renault (1998, Proposition 2.4)). In particular, if y_1 is an element of S_1 and y_2 is an element of S_2, Granger noncausality of S_1 for S_2 implies that y_1 is never causal for y_2.

Example 3.2: Pairs of Vectors S_1, S_2 if $k_3 = 2$.
Let $y_3 = [\mathtt{y_3}, \mathtt{y_4}]'$, then 4 different scenarios are possible:

(i) $S_1 = [y_1, y_3']'$, $S_2 = [y_2]$,
(ii) $S_1 = [y_1]$, $S_2 = [y_2, y_3']'$,
(iii) $S_1 = [y_1, \mathtt{y_3}]'$, $S_2 = [y_2, \mathtt{y_4}]'$,
(iv) $S_1 = [y_1, \mathtt{y_4}]'$, $S_2 = [y_2, \mathtt{y_3}]'$.

Under each scenario, Granger noncausality of S_1 for S_2 at forecast horizon $h = 1$ is sufficient for Granger noncausality of y_1 for y_2 at all forecast horizons (see Dufour & Renault (1998, Proposition 2.4)).

If the vector of third variables y_3 is univariate, there are only two possibilities (cf. cases (i) and (ii) in Example 3.2) to split up the $k_3 = 1$ variable among the two vectors S_1, S_2. However, if y_3 is multivariate ($k_3 > 1$), there exist 2^{k_3} different pairs of vectors S_1, S_2. The concept of extended Granger causality offers an elegant way to test all these scenarios implicitly in one step.

Extended Granger Causality

The concept of extended Granger causality has been developed by Boudjellaba et al. (1992a), Bruneau & Nicolai (1992a), Dufour & Renault (1994, 1998), Lütkepohl (1993) and Renault & Szafarz (1991) and is based on linear predictability at higher forecast horizons $h \geq 1$.

Definition 3.2: Granger Noncausality up to Forecast Horizon $\tilde{h}$.
y_1 is Granger noncausal for y_2 up to forecast horizon $\tilde{h}$ if

$$\mathrm{MSE}(y_{2,t}(h)) \;=\; \mathrm{MSE}(y^{\star}_{2,t}(h)) \quad \text{for all } h = 1, \ldots, \tilde{h}.$$

If $k_2 = 1$ and

$$\mathrm{MSE}(y_{2,t}(h)) \;<\; \mathrm{MSE}(y^{\star}_{2,t}(h)) \quad \text{for at least one } h \in \{1, \ldots, \tilde{h}\},$$

then y_1 is said to cause y_2 at forecast horizon h.

If $k_2 > 1$, Granger causality of y_1 for y_2 at forecast horizon h holds if

$$\text{MSE}(y_{2,t}(h)) \neq \text{MSE}(y^{\star}_{2,t}(h))$$

for at least one $h \in \{1, \ldots, \tilde{h}\}$, and if $\text{MSE}(y^{\star}_{2,t}(h)) - \text{MSE}(y_{2,t}(h))$ is positive semidefinite.

In general, the researcher is interested in finding a causal relationship. Hence, a test of causality should have as a null hypothesis that y_1 is *never causal* for y_2. In principle, this hypothesis necessitates testing for Granger noncausality at all forecast horizons $h \to \infty$.

Fortunately, in VAR models of finite order p, this null hypothesis imposes a finite number of restrictions. Example 3.1 helps to illustrate this point: Granger noncausality of y_1 for y_2 at forecast horizons $h = 1, 2$ holds if either $\pi_{21,1} = \pi_{23,1} = 0$ and/or $\pi_{21,1} = \pi_{31,1} = 0$. Under these restrictions, it can be seen that y_1 is also noncausal for y_2 at forecast horizon $h = 3$ and indeed at all higher forecast horizons $h > 2$. The following proposition states that for a VAR model of finite order p, noncausality at a finite number of forecast horizons $\tilde{h} = pk_3 + 1$ ensures that y_1 is never Granger causal for y_2:

Proposition 3.2: (Dufour & Renault (1998, Proposition 4.5)).
If y_1 is Granger noncausal for y_2 at forecast horizons $h = 1, \ldots, \tilde{h}$ with $\tilde{h} = pk_3 + 1$, then y_1 is Granger noncausal for y_2 at all forecast horizons:

$$y_1 \overset{(h)}{\not\longrightarrow} y_2 \quad \forall\ h = 1, \ldots, pk_3 + 1 \quad \Longrightarrow \quad y_1 \overset{(\infty)}{\not\longrightarrow} y_2.$$

It follows from Proposition 3.2 that Granger noncausality of y_1 for y_2 at forecast horizon $h = 1$ implies Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ only if there are no third variables, i.e. if $k_3 = 0$. As has been pointed out before, this relation also holds if y_1 and y_2 are vectors of variables.

If $k_3 > 0$, Granger noncausality of y_1 for y_2 requires that y_1 does not directly cause y_2 at forecast horizon $h = 1$ and that y_1 does not indirectly cause y_2 at forecast horizons $h = 2, \ldots, pk_3 + 1$. The number of indirect causal chains, which have to be checked to make sure that y_1 never causes y_2, grows with the lag order p and the number of third

variables k_3 in the VAR model. The following example helps to illustrate Proposition 3.2.

Example 3.3: VAR(p) Model with $k_3 = 2$.

Let $y_t = [y_{1,t},\, y_{2,t},\, y'_{3,t}]'$ with $y'_{3,t} = [\mathtt{y_{3,t}},\ \mathtt{y_{4,t}}]$ a (1×2) vector of third variables and let y_t be generated by the following VAR model:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} \pi_{11,p} & 0 & 0 \\ 0 & \pi_{22,p} & \pi_{23,p} \\ \pi_{31,p} & \pi_{32,p} & \pi_{33,p} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ y_{3,t-p} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \tag{3.6}$$

respectively

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \mathtt{y_{3,t}} \\ \mathtt{y_{4,t}} \end{bmatrix} = \begin{bmatrix} \pi_{11,p} & 0 & 0 & 0 \\ 0 & \pi_{22,p} & \pi_{23,\mathtt{p}} & 0 \\ 0 & 0 & \pi_{33,\mathtt{p}} & \pi_{34,\mathtt{p}} \\ \pi_{41,\mathtt{p}} & 0 & 0 & \pi_{44,\mathtt{p}} \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ \mathtt{y_{3,t-p}} \\ \mathtt{y_{4,t-p}} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \mathtt{u_{3,t}} \\ \mathtt{u_{4,t}} \end{bmatrix}. \tag{3.7}$$

Model (3.7) represents a *worst case* scenario insofar as one variable can directly influence another one only with the longest possible delay of p periods. Moreover, direct Granger causality of y_1 for y_2 is excluded. An indirect influence of y_1 onto y_2 is only possible via variable $\mathtt{y_4}$. However, $\mathtt{y_4}$ only indirectly Granger causes y_2 via variable $\mathtt{y_3}$. This setup allows us to study the longest possible causal chain which may establish indirect Granger causality of y_1 for y_2. Consider therefore the h–step ahead forecast of $y_{2,t}$:

For $h \leq p$, the forecast

$$y_{2,t}(h) = \pi_{22,p}\, y_{2,t+h-p} + \pi_{23,\mathtt{p}}\, \mathtt{y_{3,t+h-p}} \tag{3.8}$$

can be based solely on the information in the past and present of $y_{2,t}$ and $\mathtt{y_{3,t}}$.

For $h > p$, the variables $y_{2,t+h-p}$ and $\mathtt{y_{3,t+h-p}}$ have to be replaced by their forecasts. This yields

$$\begin{aligned} y_{2,t}(h) &= \pi_{22,p}\pi_{22,p}\, y_{2,t+h-2p} + \pi_{22,p}\pi_{23,\mathtt{p}}\, \mathtt{y_{3,t+h-2p}} \\ &+ \pi_{23,\mathtt{p}}\pi_{33,\mathtt{p}}\, \mathtt{y_{3,t+h-2p}} + \pi_{23,\mathtt{p}}\pi_{34,\mathtt{p}}\, \mathtt{y_{4,t+h-2p}}. \end{aligned}$$

Hence for $p \leq h \leq 2p$, the forecast of $y_{2,t+h}$ can be based solely on the information in the past and present of $y_{2,t}$, $\mathtt{y_{3,t}}$ and $\mathtt{y_{4,t}}$.

For $h \geq 2p+1$, the variables $y_{2,t+h-2p}$, $\mathtt{y_{3,t+h-2p}}$ and $\mathtt{y_{4,t+h-2p}}$ have to be replaced by forecasts which yields

$$\begin{aligned} y_{2,t}(h) &= \pi_{22,p}\pi_{22,p}\pi_{22,p}\, y_{2,t+h-3p} + \pi_{22,p}\pi_{22,p}\pi_{23,\mathtt{p}}\, \mathtt{y_{3,t+h-3p}} \\ &+ \pi_{22,p}\pi_{23,\mathtt{p}}\pi_{33,\mathtt{p}}\, \mathtt{y_{3,t+h-3p}} + \pi_{22,p}\pi_{23,\mathtt{p}}\pi_{34,\mathtt{p}}\, \mathtt{y_{4,t+h-3p}} \\ &+ \pi_{23,\mathtt{p}}\pi_{33,\mathtt{p}}\pi_{33,\mathtt{p}}\, \mathtt{y_{3,t+h-3p}} + \pi_{23,\mathtt{p}}\pi_{33,\mathtt{p}}\pi_{34,\mathtt{p}}\, \mathtt{y_{4,t+h-3p}} \\ &+ \pi_{23,\mathtt{p}}\pi_{34,\mathtt{p}}\pi_{41,\mathtt{p}}\, y_{1,t+h-3p} + \pi_{23,\mathtt{p}}\pi_{34,\mathtt{p}}\pi_{44,\mathtt{p}}\, \mathtt{y_{4,t+h-3p}}. \end{aligned}$$

Hence, from $h = 2p+1$ onwards, the information in $y_{1,t+h-3p}$ has to be used. If the information in the past and present of $y_{1,t}$ plays any role in forecasting $y_{2,t+h}$, this should show up *at the latest* at $h = pk_3 + 1$. Since model (3.7) represents the worst case scenario, the forecast horizon $\tilde{h} = k_3 p + 1$ is the maximal forecast horizon up to which one has to check for noncausality.

Granger Noncausality at all Forecast Horizons in a VAR(p) Model

The analysis of extended Granger causality is based on the h–step ahead forecast of a variable. For VAR(1) models, these forecasts can be easily obtained using the prediction formula (3.5) on page 27. For VAR(p) models with $p > 1$, forecasts can be computed according to the same formula if the VAR(p) model is written as a VAR(1) model in companion form (cf. (2.17) on page 14):

$$Y_{t+h} = \mathbf{\Pi}^h Y_t + \sum_{s=0}^{h-1} \mathbf{\Pi}^s U_{t+h-s}. \tag{3.9}$$

Left–multiplying (3.9) with the $(k \times kp)$ matrix $J = [I_k, 0, \ldots, 0]$ yields the following representation:

$$y_{t+h} = J\mathbf{\Pi}^h Y_t + \sum_{s=0}^{h-1} J\mathbf{\Pi}^s U_{t+h-s}, \tag{3.10}$$

where $y_{t+h} = JY_{t+h}$. The first term on the right–hand side of (3.10) is the h–step ahead forecast of y_t, based on the information in $y_{t-i+1}, i \geq 1$, i.e.

$$y_t(h) = J\mathbf{\Pi}^h Y_t. \tag{3.11}$$

The second term on the right–hand side of (3.10),

$$u_t(h) = \sum_{s=0}^{h-1} J\mathbf{\Pi}^s U_{t+h-s}, \tag{3.12}$$

denotes the h–step ahead prediction error (see Lütkepohl (1991, p. 32)).

In the following,

$$\Pi^{(h)} = J\mathbf{\Pi}^h, \tag{3.13}$$

denotes the $(k \times kp)$ coefficient matrix obtained from the first k rows of $\mathbf{\Pi}^h$. This matrix can be partitioned congruent with the vector $Y_t = [y_t', y_{t-1}', \ldots, y_{t-p+1}']'$ into p $(k \times k)$ submatrices

$$\Pi^{(h)} = \left[\Pi_1^{(h)}, \Pi_2^{(h)}, \ldots, \Pi_p^{(h)}\right]. \tag{3.14}$$

Each submatrix $\Pi_i^{(h)}$ measures the marginal influence of the vector $y_{t-i+1},\ i = 1, \ldots, p$, in the conditional forecast of y_{t+h}.

Moreover, in accordance with the partition of the $(k \times 1)$ vector $y_t = [y_{1,t}, y_{2,t}, y_{3,t}']'$, each submatrix $\Pi_i^{(h)}$ can be further partitioned into

$$\Pi_i^{(h)} = \begin{bmatrix} \pi_{11,i}^{(h)} & \pi_{12,i}^{(h)} & \pi_{13,i}^{(h)} \\ \pi_{21,i}^{(h)} & \pi_{22,i}^{(h)} & \pi_{23,i}^{(h)} \\ \pi_{31,i}^{(h)} & \pi_{32,i}^{(h)} & \pi_{33,i}^{(h)} \end{bmatrix}. \tag{3.15}$$

In general, $\pi_{fg,i}^{(h)}$ denotes a $(k_f \times k_g)$ coefficient matrix which reflects the marginal influence of $y_{g,t-i+1}$ in the h–step ahead forecast of $y_{f,t}$. However, $k_1 = k_2 = 1$ if not mentioned differently. Hence, $\pi_{21,i}^{(h)}$ is a coefficient which measures the marginal influence of $y_{1,t-i+1}$ onto $y_{2,t}(h)$ while for $k_3 > 1$ $\pi_{23,i}^{(h)}$ is a $(1 \times k_3)$ vector of coefficients which measure the marginal influence of the k_3 variables in $y_{3,t-i+1}$ onto $y_{2,t}(h)$.

If $y_{1,t-i+1}$ does not improve the h–step ahead forecast of $y_{2,t}$, the corresponding coefficients $\pi_{21,i}^{(h)},\ i = 1, \ldots, p$, should be zero. This leads to the following corollary of Proposition 3.2:

Corollary 3.2: (Dufour & Renault (1998, Proposition 4.5)).

y_1 is Granger noncausal for y_2 at all forecast horizons $h \geq 1$ iff

$$\pi_{21,i}^{(h)} = 0 \quad \forall\ h = 1, \ldots, \tilde{h} \text{ and } i = 1, \ldots, p, \tag{3.16}$$

where $\tilde{h} = pk_3 + 1$.

y_1 is said to Granger cause y_2 at forecast horizon h, iff

$$\pi_{21,i}^{(h)} \neq 0$$

for at least one $i \in \{1, \ldots, p\}$.

It follows from Corollary 3.2, that Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ imposes $\tilde{h}p$ restrictions on the VAR coefficients. Moreover, these restrictions are linked by the following recursion formula (see Dufour & Renault (1998, Lemma 3.2)):

$$\pi_{21,i}^{(h)} = \pi_{21,i+1}^{(h-1)} + \pi_{21,1}^{(h-1)}\pi_{11,i} + \pi_{22,1}^{(h)}\pi_{21,i} + \pi_{23,1}^{(h-1)}\pi_{31,i}. \tag{3.17}$$

The recursion formula shows that $\pi_{21,i}^{(h)}$ consists of sums of products of the vector autoregressive coefficients.

More Insight into the Extended Granger Causality Concept

Corollary 3.3: (Dufour & Renault (1998, Lemma 3.2 and Theorem 3.2)).
With reference to (3.17), the restrictions for Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ can be written alternatively as

$$\begin{aligned} &\pi_{21,i} = 0 \quad \text{for} \quad i = 1, \ldots, p, \\ &\qquad \text{and} \\ &\pi_{23,1}^{(h)}\pi_{31,i} = 0 \quad \text{for} \quad i = 1, \ldots, p, \quad \text{and} \quad h = 1, \ldots, \tilde{h} - 1. \end{aligned} \tag{3.18}$$

Corollary 3.3 states, that the $\tilde{h}p$ restrictions for Granger noncausality at all forecast horizons can be decomposed into p linear restrictions ($\pi_{21,1} = \ldots = \pi_{21,p} = 0$) which exclude any direct causal influence of y_1 on y_2, and into $(\tilde{h} - 1)p = k_3p^2$ nonlinear restrictions ($\pi_{23,1}^{(h)}\pi_{31,i} = 0$ for $i = 1, \ldots, p$ and $h = 1, \ldots, \tilde{h} - 1$) which exclude any indirect causal influence of y_1 onto y_2.

Corollary 3.4: (Dufour & Renault (1998, Corollary 3.5)).
If the vector of third variables is univariate ($k_3 = 1$), then either set of restrictions

$$\pi_{21,i} = \pi_{23,i} = 0 \quad \forall\, i = 1, \ldots, p, \tag{3.19}$$

or

$$\pi_{21,i} = \pi_{31,i} = 0 \quad \forall\, i = 1, \ldots, p, \tag{3.20}$$

is necessary and sufficient for Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$. (See also Hsiao (1982) and Triacca (2000)).
If the vector of third variables is multivariate ($k_3 > 1$), the restrictions are only sufficient.

Example 3.1 has illustrated that if y_3 is scalar ($k_3 = 1$), there is only one Granger causal chain running from y_1 to y_3 to y_2. This chain is interrupted in either case (3.19) or (3.20).
The case of a multivariate vector of third variables $y_{3,t}$ has been studied in Example 3.2: the latter example shows that for $k_3 = 2$, there exist 4 cases how the k–dimensional vector y_t can be decomposed into two subvectors $S_{1,t}$, $S_{2,t}$ such that standard Granger noncausality of S_1 for S_2 implies that y_1 is never Granger causal for y_2. Corollary 3.4 only describes two of these four cases (i.e. cases (i) and (ii), but not cases (iii) and (iv)). This illustrates that the restrictions given in Corollary 3.4 are only sufficient but not necessary if $k_3 > 1$.
In contrast, the restrictions given in Corollary 3.3 are necessary *and* sufficient. Example 3.4 shows that these restrictions cover indeed all 4 cases of Example 3.2 .

Example 3.4: VAR(1) Model with $k_3 = 2$.
Let $y_t = (y_{1,t}, y_{2,t}, y'_{3,t})'$ with $y_{3,t} = [\mathtt{y}_{3,\mathtt{t}}, \mathtt{y}_{4,\mathtt{t}}]'$ a (2×1) vector of third variables and let y_t be generated by the following VAR model

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \mathtt{y}_{3,\mathtt{t}} \\ \mathtt{y}_{4,\mathtt{t}} \end{bmatrix} = \begin{bmatrix} \pi_{11,1} & \pi_{12,1} & \pi_{13,1} & \pi_{14,1} \\ \pi_{21,1} & \pi_{22,1} & \pi_{23,1} & \pi_{24,1} \\ & \pi_{32,1} & \pi_{33,1} & \pi_{34,1} \\ \pi_{41,1} & \pi_{42,1} & \pi_{43,1} & \pi_{44,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ \mathtt{y}_{3,\mathtt{t}-1} \\ \mathtt{y}_{4,\mathtt{t}-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \mathtt{u}_{3,\mathtt{t}} \\ \mathtt{u}_{4,\mathtt{t}} \end{bmatrix}.$$

According to Proposition 3.2, Granger noncausality of y_1 for y_2 at all forecast horizons holds if $\pi_{21,1}^{(h)} = 0$ for $h = 1, 2, 3$. This yields the following set of restrictions:

$$\begin{bmatrix} \pi_{21,1} \\ \pi_{21,1}^{(2)} \\ \pi_{21,1}^{(3)} \end{bmatrix} = \begin{bmatrix} \pi_{21,1} \\ \pi_{23,1}\pi_{31,1} + \pi_{24,1}\pi_{41,1} \\ \pi_{23,1}\pi_{33,1}\pi_{31,1} + \pi_{24,1}\pi_{43,1}\pi_{31,1} \\ +\pi_{23,1}\pi_{34,1}\pi_{41,1} + \pi_{24,1}\pi_{44,1}\pi_{41,1} \end{bmatrix} = 0.$$

Hence, there are $p = 1$ linear restriction and $p^2 k_3 = 2$ nonlinear restrictions. For instance, the second restriction is fulfilled if either

(i) $\pi_{23,1} = [\pi_{23,1}, \pi_{24,1}] = 0$ or

(ii) $\pi_{31,1} = [\pi_{31,1}, \pi_{41,1}]' = 0$ or

(iii) $\pi_{23,1} = \pi_{41,1} = 0$ or

(iv) $\pi_{24,1} = \pi_{31,1} = 0$ or

(v) $\pi_{23,1}\pi_{31,1} = -\pi_{24,1}\pi_{41,1}$ and $\pi_{24}\pi_{41,1} \neq 0$, $\pi_{23,1}\pi_{31,1} \neq 0$.

Three interesting insights can be gained:

First, if condition (i) or condition (ii) holds, the restriction at forecast horizon $h = 3$ is fulfilled automatically and is thus redundant. This problem will be addressed in Chapter 4.

Second, if condition (iii) holds, the possibility still exists that $y_1 \xrightarrow{1} \mathtt{y}_3 \xrightarrow{1} \mathtt{y}_4 \xrightarrow{1} y_2$. Alternatively, condition (iv) does not exclude that $y_1 \xrightarrow{1} \mathtt{y}_4 \xrightarrow{1} \mathtt{y}_3 \xrightarrow{1} y_2$. To explicitly exclude these Granger causal chains, the restriction at forecast horizon $h = 3$ is needed which checks whether $\pi_{34,1} = 0$ or $\pi_{43,1} = 0$. Hence, if condition (iii) or (iv) holds, all $\tilde{h}p = 3$ restrictions are needed to establish a separation as described in Example 3.2, case (iii) or (iv).

Third, if condition (v) holds, y_1 has an indirect effect onto y_2 via $\mathtt{y}_3$ and $\mathtt{y}_4$. One might thus call y_1 causal for y_2. However, under the extended Granger causality concept, y_1 is not causal for y_2 at forecast horizon $h = 2$: the indirect causal effects cancel out so that there is no prediction improvement and hence no Granger causality at this forecast horizon.[2] This distinguishes the concept of extended Granger causality from other causality concepts, which would conclude causality under this condition (see e.g. Hsiao (1982), Triacca (2000)[3]).

---

[2] If condition (v) holds, S_1 is standard Granger causal for S_2 in Example 3.2 yet y_1 is never Granger causal for y_2 according to Corollary 3.3: this explains why standard Granger noncausality of S_1 for S_2 is only a sufficient but not a necessary condition for Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$.

[3] Triacca (2000) defines y_1 to be *Hsiao noncausal* for y_2 if and only if condition (3.19) or (3.20) holds. Hence, for the present example y_1 is Hsiao noncausal for y_2 under conditions (i) and (ii) but Hsiao causal under conditions (iii) to (v).

3.1.2 Impulse Response Analysis

The concept of impulse response analysis has been made popular by Sims (1980, 1981) and can be used to study causality between variables. It defines a variable y_1 to be causal for another variable y_2 if new (unpredicted) information in $y_{1,t}$ leads us to change our forecast of $y_{2,t+h}$ for at least one $h \geq 1$ (see Hamilton (1994, p. 319)). The new information is commonly modeled as a one–time exogenous shock which occurs in one variable of the system, say y_1. If variable y_1 is causal for another variable y_2, the one–time shock in y_1 should change the path of y_2. In other words, y_2 should *respond* to the impulse in y_1, henceforth the name *impulse response analysis*.

Noncausality in Terms of Standard Impulse Response Analysis

Let

$$y_t = \sum_{i=1}^{p} \Pi_i \, y_{t-i} + u_t, \quad u_t \sim i.i.d.(0, \Sigma_u),$$

be a stable, stationary VAR(p) model. Since the innovations in u_t can be interpreted as one–step ahead forecast errors, unpredicted information in variable $y_{1,t}$ can be modeled for example as a one–unit shock in $u_{1,t}$. To see how such a shock affects the variables in y_t in subsequent periods, given that no other shocks affect the system, it is useful to consider the moving–average (MA) representation

$$y_t = \sum_{j=0}^{\infty} \Phi_j \, u_{t-j}. \tag{3.21}$$

The moving average polynomial $\Phi(L) = \sum_{j=0}^{\infty} \Phi_j L^j$ can be obtained from inversion of the vector autoregressive polynomial $\Pi(L) = (I_k - \sum_{i=1}^{p} \Pi_i L^i)$. The condition $\Phi(L)\,\Pi(L) = I_k$ yields the recursion formula (2.10) on page 11.

If the VAR(p) model is written as a VAR(1) model in companion form

$$Y_t = \mathbf{\Pi} \, Y_{t-1} + U_t,$$

the following MA representation

$$Y_t = \sum_{j=0}^{\infty} \mathbf{\Pi}^j \, U_{t-j}$$

is obtained. Left–multiplying with the $(k \times kp)$ matrix $J = [I_k\, 0 \ldots 0]$ then yields the MA representation (3.21) for y_t as well as an alternative expression for the moving average coefficient matrices:

$$\Phi_j = J\mathbf{\Pi}^j J' = \Pi_1^{(j)} \text{ and } \Phi_0 = I_k. \tag{3.22}$$

The MA coefficient matrices $\Phi_j, j > 0$, can be partitioned in accordance with the vector y_t as

$$\Phi_j = \begin{bmatrix} \phi_{11,j} & \phi_{12,j} & \phi_{13,j} \\ \phi_{21,j} & \phi_{22,j} & \phi_{23,j} \\ \phi_{31,j} & \phi_{32,j} & \phi_{33,j} \end{bmatrix}. \tag{3.23}$$

In particular, $\phi_{fg,j}$ can be interpreted as response of $y_{f,t+j}$ to a one–time, one–unit shock in $y_{g,t}$ (see Lütkepohl (1991, Chapter 2)). Assume therefore, that prior to time t all variables in y_t take on their mean value of zero. In period t, a one–time, one–unit shock occurs in the 1–step ahead forecast error of $y_{1,t}$ only, i.e. $u_{1,t} = 1$ while $u_{2,t} = u_{3,t} = 0$. To analyze how this shock affects future values of y_t, given that no other shocks occur, u_s is restricted to zero for $s \neq t$. This yields

$$y_{t+h} = \sum_{j=0}^{\infty} \Phi_j\, u_{t+h-j} = \Phi_h\, u_t = \begin{bmatrix} \phi_{11,h} \\ \phi_{21,h} \\ \phi_{31,h} \end{bmatrix}.$$

If $\phi_{21,h} \neq 0$, then a one–time one–unit shock in $y_{1,t}$ changes the h–step ahead forecast of $y_{2,t}$ and y_1 may be called causal for y_2 in terms of impulse response analysis. If $\phi_{21,h} = 0$, then $y_{2,t+h}$ does not respond to an impulse in $y_{1,t}$ and y_1 may be called non-causal for y_2 at forecast horizon h in terms of impulse response analysis. If the latter condition holds for all forecast horizons $h \geq 1$, then y_1 may be called never causal for y_2 in terms of impulse response analysis. This is summarized in the following definition:

Definition 3.3: Noncausality in Terms of Impulse Response Analysis.

If $\phi_{21,j} = 0$ for all $j \geq 1$, then y_1 is called never causal for y_2 in terms of impulse response analysis.

If $\phi_{21,j} \neq 0$ for at least one $j \geq 1$, then y_1 is called causal for y_2 in terms of impulse

response analysis at forecast horizon j.

For finite order VAR models, Proposition 3.3 states that it suffices to check only the first $p(k-1)$ responses of y_2 to an impulse in y_1 to make sure that all responses are zero.

Proposition 3.3: (Lütkepohl (1991, Proposition 3.4)).
If $\phi_{21,j} = 0$ for $j = 1, \ldots, \bar{h}$ with $\bar{h} = p(k-1)$, then y_1 is never causal for y_2 in terms of impulse response analysis.

Example 3.5: Bivariate VAR(p) Model.
Let the (2×1) vector $y_t = [y_{1,t}, y_{2,t}]'$ be generated as

$$y_t = \Pi_p y_{t-p} + u_t, \quad u_t \sim i.i.d.(0, \Sigma_u).$$

with

$$\Pi_p = \begin{bmatrix} \pi_{11,p} & \pi_{12,p} \\ \pi_{21,p} & \pi_{22,p} \end{bmatrix}$$

This model describes a worst case scenario insofar, as regressors enter only with the highest lag p.

For the present DGP, the following impulse response coefficient matrices are obtained (cf. (2.10) on page 11):

$$\Phi_j = \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix} = \begin{cases} \Pi_p^n & \forall\ j = p * n, \quad n = 1, 2, \ldots, \\ 0 & \text{otherwise.} \end{cases} \tag{3.24}$$

Since $\Phi_j = 0$ for $j < p$, a nonzero response of y_2 to an impulse in y_1 will be detected at the earliest at forecast horizon $j = p$, namely if $\phi_{21,p} = \pi_{21,p} \neq 0$. However, if $\pi_{21,p} \neq 0$, y_1 is Granger causal for y_2. This Granger causal link will be detected at forecast horizon $h = 1$. The $p - 1$ extra steps are needed because impulse response analysis only considers the (unpredicted) information in y_t, thus neglecting the information in the $p - 1$ potential regressors $y_{t-1}, \ldots, y_{t-p}$. For the present DGP, a shock in $u_{1,t}$ respectively $y_{1,t}$ is transmitted onto $y_{2,t+h}$ only after p periods:

$$y_{t+p} = \Pi_p y_t + u_{t+p} = y_t(p) + u_t(p),$$

see (3.11) and (3.12), page 32.

If there is a nonzero response of y_2 to an impulse in y_1, one might be interested in the size of this response over n periods. Therefore, the responses of $y_{2,t}(j)$ to an impulse in $y_{1,t}$ can be summed up for $j = 1, \ldots, n$ to yield the n–th interim multiplier

$$\phi_{21}(n) = \sum_{j=1}^{n} \phi_{21,j} \ . \tag{3.25}$$

Moreover, in a stable VAR model, the total effect of an impulse in y_1 onto y_2 can be measured by the total or long–run multiplier

$$\phi_{21}(\infty) = \sum_{j=1}^{\infty} \phi_{21,j} \ . \tag{3.26}$$

It follows from Definition 3.3 and Proposition 3.3, that y_1 is causal for y_2 in terms of impulse response analysis, if $\phi_{21,j} \neq 0$ for at least one $j \in \{1, \ldots, p(k-1)\}$, hence if the $p(k-1)$–th interim multiplier $\phi_{21}(p(k-1)) \neq 0$. However, the opposite does not hold: if $\phi_{21}(p(k-1)) = 0$, y_1 may nevertheless be causal for y_2 in terms of impulse response analysis since a sum may equal zero although its elements are different from zero.

Granger Causality, Extended Granger Causality and Impulse Response Analysis

In a multivariate VAR model with some third variables, y_1 may be Granger noncausal for y_2 at forecast horizon $h = 1$ and yet some responses of y_2 to a one–time one–unit impulse in y_1 are nonzero. Indeed, y_1 may be directly Granger noncausal for y_2 but Granger cause y_3 which may then Granger cause y_2. Under this scenario, a shock in y_1 will eventually lead to a response of y_2 via y_3. This indirect causality has already been illustrated in the context of extended Granger causality. Are impulse response analysis and extended Granger causality two sides of the same causality concept? The answer is "no" in general although there are DGP's where both concepts lead to the same set of noncausality restrictions.

The h–step ahead forecast of a VAR(p) model with $p \geq 1$ is (cf. (3.10) on page 32)

$$\begin{aligned} y_{t+h} &= \Pi_1^{(h)} y_t + \Pi_2^{(h)} y_{t-1} + \ldots + \Pi_p^{(h)} y_{t-p+1} + \sum_{s=0}^{h-1} \Pi_1^{(s)} u_{t+h-s} \\ &= y_t(h) + u_t(h). \end{aligned}$$

The concept of extended Granger causality analyzes the contribution of the information in $y_{1,t-i},\ i \geq 0$ to the h–step ahead forecast of $y_{2,t}$. Granger noncausality of y_1 for y_2 at all forecast horizons holds if

$$\pi_{21,1}^{(h)} = \pi_{21,2}^{(h)} = \ldots = \pi_{21,p}^{(h)} = 0 \quad \forall\ h = 1, \ldots, \tilde{h}, \quad \text{where} \quad \tilde{h} = pk_3 + 1. \tag{3.27}$$

Impulse response analysis also investigates in how far the information in $y_{1,t-i}$ changes the forecast of $y_{2,t+h}$ in subsequent forecast periods, but only for $i = 0$. Let $u_t = [1,\ 0,\ 0]'$ characterize the one–time one–unit impulse in $y_{1,t}$. This changes the h–step ahead forecast of y_t to

$$y_t(h) = \Pi_1^{(h)}(y_t + u_t) + \Pi_2^{(h)} y_{t-1} + \ldots + \Pi_p^{(h)} y_{t-p+1}.$$

For $\pi_{21,1}^{(h)} \neq 0$, the shock in $y_{1,t}$ leads to a nonzero response of $y_{2,t}$ at forecast horizon h. It follows from Proposition 3.3, that y_1 is never causal for y_2 in terms of impulse response analysis if

$$\pi_{21,1}^{(h)} = 0, \quad \forall\ h = 1, \ldots, \bar{h}, \quad \text{where} \quad \bar{h} = p(k-1). \tag{3.28}$$

What can be learned from a comparison of the noncausality restrictions under either concept?

Lemma 3.1: In the bivariate VAR model, standard Granger noncausality of y_1 for y_2 is necessary and sufficient for noncausality of y_1 for y_2 in terms of impulse response analysis.

Proof: Granger noncausality of y_1 for y_2 holds if $\pi_{21,i} = 0$ for all $i = 1, \ldots, p$ (see Proposition 3.1). In the bivariate VAR(p) model, these restrictions imply an upper triangular vector autoregressive polynomial matrix (see (3.2) on page 23). To obtain the moving average representation, the vector autoregressive polynomial matrix has to be inverted. It follows from the inversion rules for partitioned matrices, that the inverse

of a nonsingular upper triangular matrix is again upper triangular (see e.g. Lütkepohl (1996a, 9.14.1 (3))). Hence, $\pi_{21,i} = 0$ for $i = 1, \ldots, p$ implies and is implied by $\phi_{21,j} = 0$ for $j \geq 0$. Noncausality of y_1 for y_2 in terms of impulse response analysis follows from Definition 3.3 (see Sims (1972), Hosoya (1977), Lütkepohl (1991, Chapter 2)). □

Moreover, Granger noncausality at forecast horizon $h = 1$ implies Granger noncausality at all forecast horizons $h \geq 1$. Hence, in the bivariate VAR model the restriction $\pi_{21,i} = 0$ for all $i = 1, \ldots, p$ characterizes Granger noncausality at any forecast horizon $h \geq 1$ as well as noncausality in terms of impulse response analysis.

Lemma 3.2: In VAR(p) models with $k_1 = k_2 = 1$ and $k_3 > 0$, Granger noncausality at all forecast horizons $h \geq 1$ implies that all responses of y_2 to a one–time one–unit shock in y_1 are zero.

Proof: It follows from Proposition 3.2 that

$$\begin{aligned} & \pi_{21,i}^{(h)} = 0 && \forall\ h = 1, \ldots, \tilde{h}, && i = 1, \ldots, p, \\ \Rightarrow\ & \pi_{21,i}^{(h)} = 0 && \forall\ h \geq 1, && i = 1, \ldots, p, \\ \Rightarrow\ & \pi_{21,1}^{(j)} = \phi_{21,j} = 0 && \forall\ j \geq 1. \end{aligned}$$

It thus follows from Definition 3.3 that y_1 is never causal for y_2 in terms of impulse response analysis. (See also Dufour & Tessier (1992, Proposition 1), Bruneau & Nicolai (1992a), Dufour & Renault (1994, 1998)). □

Lemma 3.3: In VAR(1) models with $k_1 = k_2 = 1$ and $k_3 > 0$, zero responses of $y_{2,t}(h)$ to a one–time one–unit shock in $y_{1,t}$ for all $h \geq 1$ are necessary and sufficient for standard Granger noncausality and extended Granger noncausality at all forecast horizons.
In VAR(p) models with $k_1 = k_2 = 1$, $k_3 > 0$ and $p > 1$, zero responses of $y_{2,t}(h)$ to a one–time one–unit shock in $y_{1,t}$ for all $h \geq 1$ are necessary but not sufficient for Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$.

Proof: y_1 is never causal for y_2 in terms of impulse response analysis if (3.28) holds. For $p = 1$, the set of restrictions in (3.28) coincides with that in (3.27) and Granger

noncausality at all forecast horizons $h \geq 1$ follows.

For $p > 1$, imposing the restrictions in (3.28) still leaves open the possibility that $\pi_{21,i}^{(h)} \neq 0$ for some $i > 1$ and $h \geq 1$ and hence that (3.27) is violated. It follows from Proposition 3.3 and (3.17) on page 34 that y_1 is never causal for y_2 in terms of impulse response analysis if the following set of restrictions holds:

$$\begin{aligned}
\pi_{21,1} &= 0, \\
\pi_{21,1}^{(2)} &= \pi_{21,2} + \pi_{21,1}\pi_{11,1} + \pi_{22,1}\pi_{21,1} + \pi_{23,1}\pi_{31,1} = 0, \\
\pi_{21,1}^{(3)} &= \pi_{21,2}^{(2)} + \pi_{21,1}^{(2)}\pi_{11,1} + \pi_{22,1}^{(2)}\pi_{21,1} + \pi_{23,1}^{(2)}\pi_{31,1} = 0, \\
\pi_{21,1}^{(4)} &= \pi_{21,2}^{(3)} + \pi_{21,1}^{(3)}\pi_{11,1} + \pi_{22,1}^{(3)}\pi_{21,1} + \pi_{23,1}^{(3)}\pi_{31,1} = 0, \\
\dots &= \dots \\
\pi_{21,1}^{(\bar{h})} &= \pi_{21,2}^{(\bar{h}-1)} + \pi_{21,1}^{(\bar{h}-1)}\pi_{11,1} + \pi_{22,1}^{(\bar{h}-1)}\pi_{21,1} + \pi_{23,1}^{(\bar{h}-1)}\pi_{31,1} = 0.
\end{aligned}$$

Inserting restrictions succinctly yields

$$\begin{aligned}
\pi_{21,1} &= 0, \\
\pi_{21,2} &= -\pi_{23,1}\pi_{31,1}, \\
\pi_{21,2}^{(2)} &= -\pi_{23,1}^{(2)}\pi_{31,1}, \\
\dots &= \dots \\
\pi_{21,2}^{(\bar{h})} &= -\pi_{23,1}^{(\bar{h}-1)}\pi_{31,1}.
\end{aligned}$$

These restrictions may be fulfilled even if $\pi_{21,2}^{(h)} \neq 0$ for some $h \geq 1$. However, in the latter case, there exists Granger causality of y_1 for y_2 at forecast horizon h. (See also Theorem 3.3 of Dufour & Renault (1998)). □

Lemma 3.4: If y_1 causes y_2 in terms of impulse response analysis, then y_1 is also Granger causal for y_2 at some forecast horizon $h \geq 1$.

Proof: The proof is self–evident if a nonzero response arises at some forecast horizon $1 \leq h \leq \tilde{h}$. In this case, the restrictions in (3.27) are violated.

Assume now that a nonzero response arises the first time at some forecast horizon $h^\star$

with $\tilde{h} < h^\star \leq \bar{h}$, i.e.

$$\pi_{21,1}^{(h)} = 0 \quad h = 1, \ldots, h^\star - 1, \tag{3.29}$$

but

$$\pi_{21,1}^{(h^\star)} \neq 0. \tag{3.30}$$

In this case, the necessary and sufficient conditions of Corollary 3.2 are violated: if y_1 is Granger causal for y_2 at forecast horizon $h^\star > \tilde{h}$, this Granger causality must already show up at some forecast horizon $h \in 1, \ldots, \tilde{h}$. Hence, the latter situation can be ruled out. (See also Corollary 3.4 of Dufour & Renault (1998): a nonzero impulse response function violates the necessary conditions for Granger noncausality at all forecast horizons.) □

Lemma 3.5: In VAR(p) models with $k_1 = k_2 = 1$, $k_3 > 0$ and $p > 1$, zero responses of $y_{2,t}(h)$ to a one–time one–unit impulse in $y_{1,t}$ at all forecast horizons $h \geq 1$ plus Granger noncausality of y_1 for y_2 at forecast horizon $h = 1$ are necessary but not sufficient for Granger noncausality at all forecast horizons $h > 1$.
They are necessary and sufficient only in the special case where $k_3 = 1$.

Proof: see the proof in Bruneau & Nicolai (1992a, Theorem and Corollary 1) and in Dufour & Renault (1998, Corollary 3.5). □

Lemmas 3.2 to 3.5 illustrate that in general the restrictions for Granger noncausality of y_1 for y_2 at all forecast horizons encompass those for noncausality in terms of impulse response analysis. For this reason, some authors regard the extended Granger causality concept as a generalized version of standard impulse response analysis (see Bruneau & Nicolai (1992a), Dufour & Renault (1994, 1998)).
However, the fact that in the stable, stationary VAR case, zero impulse response coefficients can be formulated as restrictions on the VAR coefficients which in some cases mirror the restrictions for standard or extended Granger noncausality does not make the concepts alike. Indeed, the conceptual differences become obvious once more sophisticated versions of standard impulse response analysis are considered.

Noncausality in Terms of Orthogonalized Impulse Response Analysis

The characterization of noncausality in terms of impulse response analysis has been derived under the assumption of a one–unit shock in $y_{1,t}$, given that no shocks in the other variables occur. This assumption is hard to justify if in the moving average representation (3.21) on page 37 the vector of error terms u_t has a covariance matrix Σ_u with nonzero off–diagonal elements. In this case, the error terms of different equations are contemporaneously correlated. Tracing out the response of one variable to an isolated shock in another variable neglects these contemporaneous correlations and hence the sample information (Sims (1981)).

Both problems can be solved by taking into account the information in the covariance matrix Σ_u. Consider therefore a Cholesky decomposition of the covariance matrix

$$E(u_t u_t') = \Sigma_u = PP' \quad \text{with} \quad P^{-1}\Sigma_u P^{-1'} = I_k$$

and P either a lower or upper triangular matrix. For a given matrix P, the MA representation (3.21) can be written as

$$y_t = \sum_{j=0}^{\infty} \Phi_j P P^{-1} u_{t-j} = \sum_{j=0}^{\infty} \Theta_j w_{t-j}$$

with $\Theta_0 = P$, $\Theta_j = \Phi_j P$ and $w_t = P^{-1}u_t$ a vector of orthogonal residuals with covariance matrix $\Sigma_w = I_k$.

Let Θ_j be partitioned in accordance with y_t, i.e.

$$\Theta_j = \begin{bmatrix} \theta_{11,j} & \theta_{12,j} & \theta_{13,j} \\ \theta_{21,j} & \theta_{22,j} & \theta_{23,j} \\ \theta_{31,j} & \theta_{32,j} & \theta_{33,j} \end{bmatrix},$$

then $\theta_{21,j}$ measures the responses of variable $y_{2,t}$ to a shock of size one unit in innovation $w_{1,t}$, j periods ago. For $\Theta_0 \neq I_k$, $y_{2,t}$ may now instantaneously respond to an impulse in $y_{1,t}$. This leads to the following definition of noncausality:

Definition 3.4: Noncausality in Terms of Orthogonal Impulse Response Analysis.
y_1 is called noncausal for y_2 in terms of orthogonal impulse respone analysis if $\theta_{21,j} = 0$

for all $j \geq 0$.

If $\theta_{21,j} \neq 0$ for at least one $j \geq 0$, then y_1 is called causal for y_2 in terms of orthogonal impulse response analysis.

Proposition 3.3 carries over to orthogonalized impulse responses (see Lütkepohl (1991, Proposition 2.5)). Hence, it suffices to check the first $j = 0, 1, \ldots, \bar{h}$ impulse responses to make sure that all orthogonalized responses of $y_{2,t}(j)$ to an impulse in $y_{1,t}$ are zero for $j \geq 0$.

The following example illustrates how orthogonalization changes the interpretation of impulse responses:

Example 3.6: Bivariate VAR(p) Model.

Let the (2×1) vector $y_t = [y_{1,t}, y_{2,t}]'$ be generated as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix} \begin{bmatrix} u_{1,t-j} \\ u_{2,t-j} \end{bmatrix}, \tag{3.31}$$

with $u_t = [u_{1,t}, u_{2,t}]' \sim i.i.d.(0, \Sigma_u)$, $\Sigma_u = PP'$ and P a lower triangular matrix. Then (3.31) can be written as

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \sum_{j=0}^{\infty} \begin{bmatrix} \theta_{11,j} & \theta_{12,j} \\ \theta_{21,j} & \theta_{22,j} \end{bmatrix} \begin{bmatrix} w_{1,t-j} \\ w_{2,t-j} \end{bmatrix}, \tag{3.32}$$

with

$$\begin{bmatrix} \theta_{11,j} & \theta_{12,j} \\ \theta_{21,j} & \theta_{22,j} \end{bmatrix} = \begin{bmatrix} \phi_{11,j} & \phi_{12,j} \\ \phi_{21,j} & \phi_{22,j} \end{bmatrix} \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix},$$

$$\begin{bmatrix} w_{1,t} \\ w_{2,t} \end{bmatrix} = \begin{bmatrix} P^{11} & 0 \\ P^{21} & P^{22} \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix},$$

and P^{ij} denoting the ij–th element of the inverse of P.

Note, that orthogonalization suppresses the contemporaneous correlation between the innovations of different equations, i.e. $E(w_t w_t') = I_k$. However, the information on the contemporaneous correlation shows up in the matrix $\Theta_0 = P$ which in contrast to standard impulse responses allows for an instantaneous response. The coefficients in

Θ_0 are therefore sometimes called impact multipliers. For instance, the coefficient $\theta_{21,0}$ measures the response of $y_{2,t}$ to an instantaneous shock in $w_{1,t}$. Thereby, the form of the matrix P decides on the direction of the instantaneous causality. For the present example, the lower triangularity of P allows for an instantaneous causal link from $y_{1,t}$ to $y_{2,t}$ but does not allow for a feedback: for $\theta_{21,0} = P_{21} \neq 0$, a shock in $w_{1,t}$ evokes an immediate response of $y_{2,t}$. On the other hand, $\theta_{12,0} = P_{12} = 0$, so that impulses in $y_{2,t}$ can affect $y_{1,t}$ at the earliest after one period.

The decision on the direction of the instantaneous causality cannot be derived from the data. Since correlation is a temporally symmetrical concept, one cannot decide on the basis of the timing of the data whether y_1 is instantaneously causal for y_2 or vice versa. Accordingly, the Choleski decomposition of the covariance matrix Σ_u works for P either upper or lower triangular.

The instantaneous causal direction depends on the ordering of the variables in y_t and not on sample information. For instance, in Example 3.6, P is chosen lower triangular so that $w_{1,t} = u_{1,t}$ while $w_{2,t}$ is a linear combination of $u_{1,t}$, $u_{2,t}$. A one–unit shock in $w_{1,t}$ corresponds to a shock of size one standard deviation in $u_{1,t}$ but a one–unit shock in $w_{2,t}$ corresponds to shocks of size one standard deviation in $u_{1,t}$ and $u_{2,t}$ simultaneously. Hence this ordering assumes that shocks in $y_{2,t}$ must be accompanied by simultaneous shocks in $y_{1,t}$ while it is possible to have shocks in $y_{1,t}$ only. The interpretation of the impulse responses hinges on the economic plausibility of such orderings.

Sims (1980, 1981) suggests to compute and compare orthogonalized impulse response functions for all possible different causal orderings. This allows to check the influence and the plausibility of all different possible Wold causal chains. However, the influence of the prior information grows with the number of variables in the system. For instance, in a trivariate VAR model there are $3! = 6$ possible causal orderings. The larger the number of variables, the more difficult it will be to decide for a specific causal order on economic commonsense.

There exist also decompositions of $\Sigma_u = PP'$ which yield a nontriangular matrix P and hence allow for a feedback situation. Such a decomposition may be chosen if economic theory or prior information support a model where one–unit shocks in the orthogonalized errors correspond to simultaneous shocks in all original innovations $u_{1,t}, \ldots, u_{k,t}$.

Forecast error variance decomposition is a means to summarize the information in the orthogonal impulse responses: if y_1 is causal for y_2, the innovations in $y_{1,t}$ should explain a significant proportion of the forecast error variance of $y_{2,t}$ at forecast horizon h. To gain an idea in how far innovations in $y_{1,t}$ can explain the mean–squared error of the h–step ahead forecast of $y_{2,t}$, the proportion of $MSE(y_{2,t}(h))$, accounted for by innovations in $y_{1,t}$, can be computed. Summing up these proportions for all forecast horizons $h \geq 1$ yields insight into the long–run relationship between y_1 and y_2. This procedure is called *innovation accounting* (see Sims (1980, 1981)).

Orthogonalized impulse response analysis and innovation accounting are based on prior information about instantaneous causal directions while Granger causality tests are not. Due to the different underlying information sets, the duality between Granger noncausality and zero impulse responses which arises in the bivariate VAR model does not hold with orthogonalized impulse responses.

Granger Causality and Orthogonalized Impulse Response Analysis

In contrast to standard impulse response functions, orthogonalized impulse responses depend not only on the sample information but also on the assumptions about the instantaneous causal links as reflected in the choice of the matrix P and the ordering of the variables in y_t. As consequence, noncausality of y_1 for y_2 in terms of (standard) impulse response analysis neither implies nor is implied by noncausality of y_1 for y_2 in terms of orthogonalized impulse response analysis. Moreover, the relationship between Granger noncausality and noncausality in terms of impulse response analysis which holds in bivariate VAR(p) models, does not hold with orthogonal impulse responses in general:

It follows from Lemma 3.1 that restricting $\phi_{21,j} = 0$ for $j = 1, 2, \ldots$ in (3.31) implies and is implied by Granger noncausality of y_1 for y_2. On the other hand, y_1 may be Granger noncausal for y_2 and yet in (3.32) the orthogonalized impulse responses $\theta_{21,j} = \phi_{21,j} P_{11} + \phi_{22,j} P_{21} = \phi_{22,j} P_{21} \neq 0$ for some $j \geq 0$. The orthogonalized impulse responses $\theta_{21,j} = 0$ only under the additional condition that y_1 is not instantaneously causal for y_2, hence if $P_{21} = 0$.

On the other hand, since P is chosen lower triangular in Example 3.6, $P_{12} = 0$ so that $\theta_{12,j} = \phi_{12,j} P_{22} = 0$ in (3.32) if and only if $\phi_{12,j} = 0$. But $\theta_{12,j}$ measures the response of $y_{1,t+j}$ to an impulse in $w_{2,t} = P^{21}u_{1,t} + P^{22}u_{2,t}$ and does not coincide with the standard impulse response coefficient $\phi_{12,j}$ which measures the response of $y_{1,t+j}$ to an impulse in $u_{2,t}$ only. Again, both concepts only coincide for $P^{21} = 0$, hence if there is no instantaneous correlation between $y_{1,t}$ and $y_{2,t}$.
Since instantaneous causality does not play a role in Granger causality and standard impulse response analysis, these concepts will not be considered here. Moreover, the examples illustrate that Lemmas 3.1 to 3.5 do not carry over to orthogonalized impulse response analysis.

Concepts which allow for instantaneous causality or explicitly model instantaneous relationships are for example orthogonal impulse response analysis, variance decomposition or innovation accounting, generalized impulse response analysis and structural VAR model analysis. The interested reader is referred to Judge et al. (1988, Chapter 18), Lütkepohl (1991, Chapter 2), Hamilton (1994, Chapter 11), Amisano & Giannini (1997), Breitung (1998) and Pesaran & Shin (1998).

3.2 Causality in Nonstationary VAR Models

It is now assumed that some or all of the k variables in y_t are integrated of order one, but that y_t still admits a finite order levels VAR(p) representation

$$\Pi(L)\, y_t \;=\; u_t,$$

where $u_t \sim$ i.i.d.$(0, \Sigma_u)$ with Σ_u a nonsingular covariance matrix and $\Pi(L) = I_k - \sum_{i=1}^{p} \Pi_i$. In contrast to the stable, stationary VAR(p) model, the determinantal polynomial $det(\Pi(z))$ now has roots on the unit circle. In the following, it will be illustrated in how far the restrictions for Granger noncausality, extended Granger noncausality and noncausality in terms of impulse response analysis carry over to the nonstationary VAR case. Moreover, a brief summary of related concepts will be given.

3.2.1 Standard and Extended Granger Causality

In the stationary VAR model, the definitions of Granger causality and extended Granger causality have been based on the optimal (linear minimum MSE) forecast of y_{t+h} for $h \geq 1$:

$$y_t(h) \;=\; \Pi^{(h)}\, Y_t = J\mathbf{\Pi}^h\, Y_t.$$

This prediction formula still yields minimum (linear) MSE forecasts in the nonstationary VAR case, so that Granger noncausality at forecast horizon $h \geq 1$ is characterized by the same set of restrictions in the nonstationary case as in the stable, stationary case (Granger (1988)). These restrictions are given in Propositions 3.1 and 3.2.

In contrast to stationary VAR models, nonstationary VAR models allow a joint analysis of variables with different statistical features:
Stationary variables have a finite variance and fluctuate around their mean. The optimal long range forecast of a stationary process is therefore its process mean (Lütkepohl (1991, p. 32)). As consequence, the information in the past of a stationary variable $y_{1,t}$ can help to improve the h–step ahead forecast of another stationary variable $y_{2,t}$ in particular for h small. Hence, Granger causality between stationary variables is related to short–run forecastability (Granger (1988)).

An $I(1)$ variable has a variance which increases with time. As consequence, long range forecasts of integrated variables are beset with forecast uncertainty which increases with the forecast horizon h. Moreover, two $I(1)$ series may drift apart unless they are tied together by a cointegration relation. Cointegration thus represents a comovement of time series in the long–run. If two $I(1)$ variables are cointegrated, there must be Granger causality between these variables in at least one direction (see Engle & Granger (1987), Granger (1987, Theorem 3)). In this case the Granger causal variable helps to predict the other variable in the long–run.

Despite their different features, there may also exist Granger causality between a stationary variable and an $I(1)$ variable: let $y_{1,t}$ be a stationary variable and let $y_{2,t} \sim I(1)$, then $y_{1,t}$ may Granger cause the stationary variable $\Delta y_{2,t}$ and hence indirectly also $y_{2,t} = y_{2,t-1} + \Delta y_{2,t}$. In this case, $y_{1,t}$ is said to Granger cause $y_{2,t}$ in the short–run.

Standard Granger Causality in the Error Correction Representation

The error correction representation allows for a joint analysis of stationary and integrated variables and helps to illustrate short–run and long–run causal relationships. Let

$$y_t = \Pi_1 \, y_{t-1} + \ldots + \Pi_p \, y_{t-p} + u_t, \tag{3.33}$$

be generated by a VAR(p) model, then the error correction representation is

$$\Delta y_t = -\Pi(1) \, y_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \, \Delta y_{t-j} + u_t, \tag{3.34}$$

with $\Pi(1)$ a $(k \times k)$ matrix of rank $0 \leq r < k$. Furthermore, let $\Pi(1)$ and Γ_j be partitioned in accordance with y_t, i.e.

$$\Pi(1) = \begin{bmatrix} \Pi(1)_{11} & \Pi(1)_{12} & \Pi(1)_{13} \\ \Pi(1)_{21} & \Pi(1)_{22} & \Pi(1)_{23} \\ \Pi(1)_{31} & \Pi(1)_{32} & \Pi(1)_{33} \end{bmatrix} \quad \text{and} \quad \Gamma_j = \begin{bmatrix} \gamma_{11,j} & \gamma_{12,j} & \gamma_{13,j} \\ \gamma_{21,j} & \gamma_{22,j} & \gamma_{23,j} \\ \gamma_{31,j} & \gamma_{32,j} & \gamma_{33,j} \end{bmatrix}.$$

Standard Granger noncausality of y_1 for y_2 is characterized by exclusion restrictions on the regressors $y_{1,t-1}, \ldots, y_{1,t-p}$ in the $y_{2,t}$–equation. In the error correction representation (3.34), these restrictions translate into exclusion restrictions on the regressors $y_{1,t-1}$ and $\Delta y_{1,t-i}$, $i = 1, \ldots, p-1$. The restrictions which characterize Granger noncausality of y_1 for y_2 in the error correction representation are given in the following

proposition:

Proposition 3.4: Granger Noncausality at Forecast Horizon $h = 1$.
y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$ if

$$\Pi(1)_{21} = 0 \quad \text{and} \quad \gamma_{21,j} = 0 \quad \forall\, j = 1, \ldots, p-1. \tag{3.35}$$

If condition (3.35) does not hold, then y_1 is called Granger causal for y_2 at forecast horizon $h = 1$.
If all variables in y_t are integrated of order one but not cointegrated, $\Pi(1)$ equals the null matrix and the error correction model reduces to a VAR model in first differences. In this case, condition (3.35) only requires that $\gamma_{21,j} = 0$ for all $j = 1, \ldots, p-1$.

If $\text{rk}(\Pi(1)) = r$ with $0 < r < k$, there exists a decomposition $\Pi(1) = A\,B'$ with A, B two nonzero $(k \times r)$ matrices: $A = [a_{ls}]$ and $B = [b_{ls}]$ where $l = 1, \ldots, k$; $s = 1, \ldots, r$. Standard Granger noncausality of y_1 for y_2 requires that $\Pi(1)_{21} = a_{21}b_{11} + \ldots + a_{2r}b_{r1} = 0$. Hence, in the error correction representation Granger noncausality at forecast horizon $h = 1$ imposes nonlinear restrictions on the EC coefficients.

Long–Run and Short–Run Granger Causality

If the variables in y_t are integrated of order one and also cointegrated, $\Pi(1)$ is a nonzero matrix of rank $0 < r < k$. Moreover, $\Pi(1)$ contains the coefficients associated with the $I(1)$ vector y_{t-1} while the coefficients in Γ_j measure the influence of the stationary variables in Δy_{t-j}. Based on the different stochastic properties of the regressors, $\Pi(1)$ is sometimes called the *long–run* matrix and the restriction $\Pi(1)_{21} = 0$ is then interpreted as *long–run Granger noncausality*. On the other hand, the condition $\gamma_{21,j} = 0$ for $j = 1, \ldots, p-1$ characterizes *short–run Granger noncausality*. Some authors simply speak of *long–run* (*long–term*) and *short–run noncausality* (see e.g. Granger (1988), Toda & Phillips (1994), Hylleberg & Mizon (1989), Konishi & Granger (1992)). In the literature much attention has been drawn towards *long–run Granger causality*. Therefore, the concept will be discussed in more detail.

Example 3.7: Bivariate VAR(p) Model.

Let $y_t = [y_{1,t}, y_{2,t}]'$ be generated by a bivariate VAR(p) model with error correction representation

$$\begin{aligned}\begin{bmatrix} \Delta y_{1,t} \\ \Delta y_{2,t} \end{bmatrix} &= - \begin{bmatrix} \Pi(1)_{11} & \Pi(1)_{12} \\ \Pi(1)_{21} & \Pi(1)_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} \\ &+ \sum_{j=1}^{p-1} \begin{bmatrix} \gamma_{11,j} & \gamma_{12,j} \\ \gamma_{21,j} & \gamma_{22,j} \end{bmatrix} \begin{bmatrix} \Delta y_{1,t-j} \\ \Delta y_{2,t-j} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix}. \end{aligned} \tag{3.36}$$

If $y_{1,t}$ and $y_{2,t}$ are both integrated of order one, the left–hand side of (3.37) contains only variables integrated of order zero while the right–hand side contains variables integrated of order one and variables integrated of order zero. Since the variance of an $I(1)$ variable goes to infinitiy as t goes to infinity, it dominates the finite variance of a stationary variable so that the sum of an $I(1)$ and an $I(0)$ variable yields an $I(1)$ variable. However, the sum on the right–hand side of equation (3.37) must yield a vector with $I(0)$ variables only. As consequence, the matrix $\Pi(1)$ must be zero unless $y_{1,t}$ and $y_{2,t}$ are cointegrated, i.e. unless there exists a linear combination $z_t = b'y_t = b_1\, y_{1,t} + b_2\, y_{2,t}$ which is stationary.

The cointegration relation is often interpreted as a long–run equilibrium relationship. In this sense, values $z_t \neq 0$ measure the deviations from the long–run equilibrium. If $y_{1,t}$, $y_{2,t}$ are cointegrated, the rank of $\Pi(1)$ equals one and $\Pi(1) = ab'$ can be partitioned into a (2×1) loading vector a and a (2×1) cointegrating vector b, i.e.

$$\begin{bmatrix} \Pi(1)_{11} & \Pi(1)_{12} \\ \Pi(1)_{21} & \Pi(1)_{22} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \begin{bmatrix} b_1 & b_2 \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 \\ a_2 b_1 & a_2 b_2 \end{bmatrix}.$$

The loading vector a must contain at least one nonzero element. As consequence, z_{t-1} and hence $y_{1,t-1}$, $y_{2,t-1}$ enter into at least one equation. If $\Pi(1)_{21} \neq 0$, the information in $y_{1,t-1}$ helps to predict $\Delta y_{2,t}$ and hence $y_{2,t} = y_{2,t-1} + \Delta y_{2,t}$. In this case y_1 is called long–run Granger causal for y_2. $\Pi(1)_{12} \neq 0$ characterizes long–run Granger causality of y_2 for y_1 and $\Pi(1)_{21} \neq 0$, $\Pi(1)_{12} \neq 0$ a feedback system.

On the other hand, nonzero values $\gamma_{21,j} \neq 0$ for some $j = 1, \ldots, p-1$ are said to describe *short–run Granger causality* as these coefficients measure the influence of the stationary variables $\Delta y_{1,t-i}$ which can have only transitory effects by definition.

Since noncausality is defined as absence of causality, $\Pi(1)_{21} = 0$ characterizes *long–run Granger noncausality* of y_1 for y_2 while *short–run Granger noncausality* of y_1 for y_2 holds if $\gamma_{21,j} = 0$ for all $j = 1, \ldots, p-1$. Since Granger noncausality restricts all coefficients $\pi_{21,i} = 0$ for $i = 1, \ldots, p$, it also restricts $\Pi(1)_{21} = \sum_{i=1}^{p} \pi_{21,i}$ to zero and hence implies long–run Granger noncausality. However, the opposite does not hold. It is therefore possible to have long–run Granger noncausality of y_1 for y_2 and yet y_1 Granger causes y_2.

In the following chapters, interest centers on tests of Granger noncausality and not specifically on tests of long–run Granger noncausality. However, since the notion *long–run causality* appears quite frequently in studies of (co–)integrated time series (see e.g. Osterberg (1992), Beeby et al. (1995), Granger & Lin (1995), Dufour & Renault (1994, 1998), Caporale et al. (1998), Bruneau & Jondeau (1999)), some further remarks are given below:

Economic theory usually has more to say about long–run relations than about the short–run dynamics. If economic theory postulates a causal direction, it might be interesting to test for Granger causality in the cointegration relations only, as these are interpreted as the long–run equilibrium relations. Indeed, if two variables $y_{1,t}$, $y_{2,t}$ are integrated of order one and cointegrated, there must be *long–run Granger causality* in at least one direction (see Granger (1986)).

Since Granger causality is measured by prediction improvement, the question arises whether long–run Granger causality also means prediction improvement in the long–run?

For nonstationary VAR models, the h–step ahead forecast of y_t can be computed just as in the stable case (cf. (3.11) on page 32), but the forecast mean squared errors of $y_{1,t+h}$, $y_{2,t+h}$ may now be unbounded for $h \to \infty$. As consequence, long–run forecasts may be afflicted with extremely large forecast uncertainty. However, if $y_{1,t}$, $y_{2,t} \sim I(1)$ and y_1 is long–run Granger causal for y_2, h–step ahead forecasts of $y_{2,t}$ which take into account the cointegration relation will have bounded mean squared errors for $h \to \infty$ (Engle & Yoo (1987), Lütkepohl (1991, Chapter 11)). In this sense, long–run Granger causality of y_1 for y_2 results in long–run prediction improvement.

Nevertheless, long–run Granger causality does not mean that long–run forecasts can be obtained from the cointegration relations alone. In general, forecasts have to be based on the whole data generating process which includes the short–run dynamics (see Lütkepohl (1994b) and Lütkepohl (1994c)).

In this thesis, a vector y_t is $I(1)$ if at least one variable in y_t is integrated of order one. The vector y_t is said to be cointegrated with cointegrating vector b_s if $b_s' y_t \sim I(0)$. Let $y_t = [y_{1,t}, y_{2,t}]'$ with $y_{1,t} \sim I(1)$ and $y_{2,t} \sim I(0)$. Assume a vector $b_1 = [0, b_{21}]'$ with b_{21} a nonzero scalar which ensures that $b_1' y_t \sim I(0)$, then there exists a cointegrating vector. However, there is no cointegrating relation which can be interpreted in the sense of a long–run economic equilibrium situation and hence there is no long–run Granger causality. The example demonstrates that long–run Granger causality has to be interpreted with care in a joint analysis of stationary and $I(1)$ variables.

Note, that the notion *long–run causality* has also been used in different contexts: for instance, Dufour & Renault (1994, 1998) use the similar notion *causality in the long run* to denote Granger causality at higher forecast horizons $h > 1$. In contrast to the former notion *long–run Granger causality*, their notion is not restricted to the relationships between $I(1)$ variables.
Granger & Lin (1995) use spectral decomposition to study causality at certain frequencies as proposed by Hosoya (1991). In this context, Granger & Lin define *causality in the long–run* as causality at very small frequencies ω with $\omega \to 0$.
Stock & Watson (1989) use the notion *neutrality* in place of *long–run Granger noncausality*. In contrast, Bruneau & Nicolai (1995, 1992b) and Bergman & Warne (1993) label a variable *neutral* or *totally neutral* for another variable if the responses of the latter to an impulse in the former sum to zero.

Extended Granger Causality in the Error Correction Representation

In principle, extended Granger causality can also be defined in the error correction representation. However, the restrictions for Granger noncausality at forecast horizon $h = 1$ are already nonlinear. Example 3.8 illustrates that the restrictions for Granger

noncausality at higher forecast horizons are more complicated than those derived in the levels VAR model.

Example 3.8: Trivariate VAR(2) model.

Let $y_t \sim I(1)$ be generated by a VAR(2) model with error correction representation

$$\Delta y_t = -\Pi(1)\, y_{t-1} + \Gamma_1\, \Delta y_{t-1} + u_t, \quad u_t \sim i.i.d.(0, \Sigma_u),$$

then the one–step ahead forecast is

$$\Delta y_t(1) = -\Pi(1)\, y_t + \Gamma_1\, \Delta y_t.$$

The restrictions for Granger noncausality at forecast horizon $h = 1$ have already been explored. At forecast horizon $h = 2$,

$$\begin{aligned} \Delta y_t(2) &= -\Pi(1)\, y_{t+1} + \Gamma_1\, \Delta y_{t+1} \\ &= (-\Pi(1) + \Pi(1)\Pi(1) - \Gamma_1\Pi(1))\, y_t + (\Gamma_1\Gamma_1 - \Pi(1)\Gamma_1)\, \Delta y_t. \end{aligned}$$

y_1 is Granger noncausal for y_2 at forecast horizon $h = 2$ if the information in $y_{1,t}$ and $\Delta y_{1,t}$ does not help to improve the prediction of $\Delta y_{2,t+2}$. Noncausality holds if the corresponding elements of the matrices $(-\Pi(1) + \Pi(1)\Pi(1) - \Gamma_1\Pi(1))$ and $(\Gamma_1\Gamma_1 - \Pi(1)\Gamma_1)$ equal zero. Since products of matrices are involved, the restrictions are nonlinear. Moreover, since products of short–run and long–run coefficients are involved, the error correction representation looses its advantage to distinguish between short–run and long–run effects.

3.2.2 Impulse Response Analysis

If some or all variables in y_t are integrated of order one, the vector autoregressive polynomial $\Pi(L)$ is no longer invertible. Therefore, y_t does not have a moving average representation as in (2.9), see page 11. However, y_{t+h} can be predicted just as in the stationary, stable VAR case, i.e.

$$y_t(h) = \Pi_1^{(h)} y_t + \Pi_2^{(h)} y_{t-1} + \ldots + \Pi_p^{(h)} y_{t-p+1}.$$

Let $\Phi_j = \Pi_1^{(j)}$ be partitioned as in (3.23), page 38, then $\phi_{21,j}$ still measures the (ceteris paribus) response of $y_{2,t+j}$ to a one–time, one–unit change in $y_{1,t}$. Hence, impulse responses may be computed just as in the stable, stationary VAR case. Definition 3.3

and Lemmas 3.1 to 3.5 therefore carry over to the nonstationary VAR case.
However, in nonstationary vector autoregressive models, the impulse response coefficients cannot be interpreted as moving average coefficients. Moreover, some elements $\phi_{kl,j}$ of the Φ_j matrices may no longer converge to zero as $j \to \infty$. In this case, the infinite sum $\phi_{kl}(\infty) = \sum_{j=0}^{\infty} \phi_{kl,j}$ would not be bounded and hence total impact multipliers would not exist (see i.e. Lütkepohl (1991, Chapter 11), Lütkepohl & Reimers (1992a) and Wolters (1991)).

Impulse response functions computed from equation (2.10), page 11, or from equation (3.22), page 38, represent only one way to analyze the system's reaction to one–time one–unit shocks. Other representations of the nonstationary VAR process allow to obtain different impulse response functions.

Standard Impulse Response Analysis for Δy_t

Although y_t does not posses a valid moving average representation, it follows from Wold's representation theorem that the zero–mean stationary vector $\Delta y_t = y_t - y_{t-1}$ admits a MA representation:

$$\Delta y_t = C(L)\, u_t, \tag{3.37}$$

with

$$C(L) = \sum_{j=0}^{\infty} C_j L^j,$$

$$C_j = \begin{bmatrix} c_{11,j} & c_{12,j} & c_{13,j} \\ c_{21,j} & c_{22,j} & c_{23,j} \\ c_{31,j} & c_{32,j} & c_{33,j} \end{bmatrix}$$

and $C_0 = I_k$. The element $c_{kl,j}$ measures the response of $\Delta y_{k,t}$ to a one–time one–unit change in the stationary error term $u_{l,t-j}$ respectively $\Delta y_{l,t-j}$ for $k, l = 1, 2, 3$.

If $c_{21,j} \neq 0$ for some $j \geq 1$, then $\Delta y_{2,t}$ responds to a shock in $\Delta y_{1,t}$, j periods ago, and hence Δy_1 may be called causal for Δy_2. If on the other hand $c_{21,j} = 0$ for all $j \geq 1$, then Δy_1 may be called never causal for Δy_2 in terms of this impulse response analysis. It follows from Proposition 2.4 of Lütkepohl (1991), that it suffices to check the first

$\bar{h} = p(k-1)$ responses to show that all responses are zero.

The element $c_{21,j} = \phi_{21,j} - \phi_{21,j-1}$ measures the difference between the responses of $y_{2,t}$ and $y_{2,t-1}$ to a shock in $\Delta y_{1,t-j}$, respectively to a unit change in $y_{1,t-j}$, given that $y_{1,t-j-1}$ remains unchanged. Taken the other way round, $\phi_{21,j} = \sum_{s=0}^{j} c_{21,s}$ measures the accumulated responses of the differenced variable $\Delta y_{2,t}$ to a one–time shock in $\Delta y_{1,t-s}$ for $s = 0, \ldots, j$ (Lütkepohl & Breitung (1997), Lütkepohl (1999)).

If Δy_1 is never causal for Δy_2 in terms of impulse response analysis, i.e. if $c_{21,j} = 0$ for all $j \geq 1$, then $\phi_{21,j} = 0$ for all $j \geq 1$ and hence the level variable y_1 is also never causal for the level variable y_2 in terms of impulse responses. The MA representation (3.37) thus allows to test for noncausality between the (nonstationary) level variables in y_t in a stationary framework.

If Δy_2 responds to a shock in Δy_1, hence if $c_{21,j} \neq 0$ for some $j \geq 1$, then the levels variable y_2 also responds to a shock in the levels variable y_1. However, since Δy_1, Δy_2 are stationary variables, the response of Δy_2 to an impulse in Δy_1 will finally taper off to zero in the long–run. Hence, the impulse response coefficients $c_{21,j}$ measure short–run causality of Δy_1 for Δy_2. On the other hand, depending on how fast $c_{21,j} \to 0$ for $j \to \infty$, the impulse response coefficients $\phi_{21,j} = \sum_{s=0}^{j} c_{21,s}$ may not taper off to zero as $j \to \infty$ and therefore show a long–run effect of y_1 onto y_2. Hence, the impulse response coefficients $\phi_{kl,j}$ reflect short–run as well as long–run causality.

Impulse Response Analysis in the Common Trends Representation

To separate long–run from short–run effects, it is convenient to consider the common trends representation for the vector of level variables:

$$y_t \;=\; C(1)\,\epsilon_t + \sum_{j=0}^{\infty} C_j^{\star}\, u_{t-j} + y_0 - C^{\star}(L)\, u_0,$$

with $\epsilon_t = \sum_{s=1}^{t} u_s$, $C(1)$, $C_j^{\star}$ defined in (2.15) on page 12 and $y_0 - C^{\star}(L)\, u_0$ the initial conditions. In particular, the matrix $C(1)$ has reduced rank $k - r$ with $0 < r < k$.

Similar to the separation into long–run and short–run Granger causality in the error correction representation, the common trends representation allows to distinguish between transitory and permanent shocks: transitory shocks are shocks in the stationary variables u_{t-j} which die out in the long–run as these variables return to their mean

value of zero. In contrast, shocks to nonstationary variables ϵ_t may have effects which do not die out in the long–run and are therefore permanent.

Example 3.9: Bivariate Common Trends Model.

Let $y_t = [y_{1,t},\, y_{2,t}]'$ with $y_{1,t},\, y_{2,t} \sim I(1)$ and cointegrated with cointegration rank $r = 1$. Then

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} C(1)_{11} & C(1)_{12} \\ C(1)_{21} & C(1)_{22} \end{bmatrix} \begin{bmatrix} \epsilon_{1,t} \\ \epsilon_{2,t} \end{bmatrix} + \sum_{j=0}^{\infty} \begin{bmatrix} c^{\star}_{11,j} & c^{\star}_{12,j} \\ c^{\star}_{21,j} & c^{\star}_{22,j} \end{bmatrix} \begin{bmatrix} u_{1,t-j} \\ u_{2,t-j} \end{bmatrix}.$$

For convenience, the initial conditions $y_0 - C^{\star}(L)\, u_0$ will be neglected as they do not play a role in the definition of transitory and permanent effects.

In the common trends representation, a one–time one–unit change in $y_{1,t}$ shows up in the stationary error term $u_{1,t}$ but also in the $I(1)$ term $\epsilon_{1,t} = \sum_{s=1}^{t} u_{1,s}$. Since shocks in stationary variables can only have transitory effects, the coefficient $c^{\star}_{21,j}$ measures "short–run causality". On the other hand, a shock in $\epsilon_{1,t}$ will have a permanent effect on $y_{2,t}$ unless $C(1)_{21} = 0$. Hence, the coefficient $C(1)_{21}$ measures "long–run causality" (see Mosconi & Giannini (1992)). The condition $C(1)_{21} \neq 0$ has also been labeled *persistent causality* of y_1 for y_2 (Bruneau & Nicolai (1992b, 1995)). In contrast, the condition $C(1)_{21} = 0$ has also been called *neutrality* of y_1 for y_2 (see Bruneau & Nicolai (1992b, 1995)), *long–run noncausality* (Mosconi & Giannini (1992)) or *total neutrality* (Bergman & Warne (1993)).

If a one–time, one–unit change in $y_{1,t}$ evokes neither transitory nor permanent responses of $y_{2,t+j}$ for all $j \geq 1$, i.e. if

$$C(1)_{21} = 0 \quad \text{and} \quad c^{\star}_{21,j} = 0, \quad \forall\, j = 1, \ldots, p(k-1),$$

then y_1 may be called noncausal for y_2 in the common trends representation.

If $C(1)_{21} \neq 0$, then the stochastic trend of y_1 drives y_2 and $\epsilon_{1,t}$ is called a common trend (see e.g. Stock & Watson (1988), Warne (1990) and Wolters (1991)). In structural VAR models, special interest refers to identifying and modeling common trends on grounds of economic theory. The interested reader is referred to Mellander et al. (1992), Bergman & Warne (1993), Amisano & Giannini (1997), Dolado et al. (2000), Brei-

tung & Heinemann (1998), Breitung (1998), Lütkepohl & Breitung (1997), King et al. (1991), Swanson & Granger (1992), Kaufmann (1996).

Outlook

Chapter 3 has illustrated that in stationary as well as nonstationary VAR models Granger noncausality at any forecast horizon $h \geq 1$ can be characterized by the same set of restrictions on the vector autoregressive parameters. Moreover, it has been shown that the restrictions for Granger noncausality at higher forecast horizons $h > 1$ involve nonlinear functions of the VAR coefficients.

Chapter 3 has also shown that in stationary as well as nonstationary VAR models noncausality in terms of impulse response analysis can be characterized by the same set of restrictions on the vector autoregressive parameters. The structure of these noncausality restrictions is similar to the structure of the restrictions for Granger noncausality at higher forecast horizons. It may even coincide in some cases, see Lemmas 3.1 and 3.3.

In the next chapter problems are illustrated which may occur with standard Wald tests of Granger noncausality at all forecast horizons $h \geq 1$ as well as with standard Wald tests of noncausality in terms of impulse response analysis. Problems arise if nonlinear restrictions under the null hypothesis may violate a regularity condition of the standard Wald test. Since the restrictions for noncausality of y_1 for y_2 in terms of impulse response analysis can be regarded as a special case of the restrictions for Granger noncausality at all forecast horizons $h \geq 1$ (see Lemma 3.2), main interest centers on Wald tests of the latter null hypothesis in subsequent chapters.

Chapter 4

Causality in VAR Models: Estimation and Testing.

The foregoing chapter has illustrated that Granger noncausality of y_1 for y_2 is characterized by linear restrictions on the VAR coefficients at forecast horizon $h = 1$ and nonlinear restrictions at higher forecast horizons $h > 1$. The present chapter discusses problems which arise with standard Wald tests of Granger noncausality at forecast horizons $h \geq 1$. In contrast to Lagrange Multiplier (LM) and Likelihood Ratio (LR) tests, Wald tests do not require estimation under the null hypothesis. Since estimation of the vector autoregressive coefficients under nonlinear restrictions can be tedious, Wald tests are often the preferred choice.

The standard Wald statistic has an asymptotic χ^2–distribution under the regularity condition that the covariance matrix of the vector of restrictions is nonsingular under the null hypothesis (see e.g. Engle (1984),Buse (1982)). However, nonlinear restrictions may violate this regularity condition and inference based on the standard Wald test may then be misleading (see e.g. Andrews (1987), Lütkepohl (1993), Boudjellaba et al. (1992a, b) and Dufour & Renault (1994, 1998)). In this case, LR and LM tests offer no alternative as they suffer from similar problems.

The regularity condition may also be violated under linear restrictions, if the covariance matrix of the vector autoregressive coefficients is singular. For instance, this problem arises with tests of standard Granger noncausality in cointegrated VAR models.

A closer look at the Wald statistic will help illustrate the problems outlined above: let π denote a vector of coefficients and let $\hat{\pi}$ be a consistent estimator of π. It is assumed that

$$\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \Sigma_{\hat{\pi}}),$$

where $\xrightarrow{d}$ denotes convergence in distribution and T is the sample size.
Let the $(r_g \times 1)$ vector $g(\pi)$ denote a set of possibly nonlinear restrictions on the coefficients in π and let $g(\hat{\pi})$ be a consistent estimator of $g(\pi)$ with covariance matrix

$$\Sigma_{g(\hat{\pi})} = \frac{\partial g(\pi)}{\partial \pi'} \Sigma_{\hat{\pi}} \frac{\partial g(\pi)'}{\partial \pi}.$$

Let $\hat{\Sigma}_{g(\hat{\pi})}$ be a consistent estimator of $\Sigma_{g(\hat{\pi})}$ and assume that $\hat{\Sigma}^{-1}_{g(\hat{\pi})}$ exists. Then the Wald statistic for a test of $H_0 : g(\pi) = 0$ against $H_1 : g(\pi) \neq 0$ is

$$\mathcal{W} = T g(\hat{\pi})' \hat{\Sigma}^{-1}_{g(\hat{\pi})} g(\hat{\pi}),$$

which has an asymptotic χ^2–distribution with degrees of freedom equal to r_g under the null hypothesis if the following regularity condition holds:

(R) The $(r_g \times r_g)$ matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi')\Sigma_{\hat{\pi}}(\partial g(\pi)'/\partial \pi)$ has full rank r_g.

The regularity condition does not hold if (i) the row rank of the Jacobian matrix $\partial g(\pi)/\partial \pi'$, evaluated at the true parameter values, is less than r_g. Furthermore, the regularity condition need not hold if (ii) the covariance matrix $\Sigma_{\hat{\pi}}$ is singular. In these cases, the Wald statistic may not have a χ^2–distribution in general (see e.g. Andrews (1987)).
It will be shown in this chapter that a case (i) violation of the regularity condition may arise with a standard Wald test that y_1 is Granger noncausal for y_2 up to some forecast horizon $h > 1$. For simplicity of exposition, case (i) will be explained in the stationary, stable VAR model framework. However, it is also relevant in nonstationary VAR models. Moreover, in nonstationary VAR models, the covariance matrix of the levels VAR coefficients $\Sigma_{\hat{\pi}}$ is singular. Consequently, a standard Wald test that y_1 does not Granger cause y_2 at some forecast horizon $h \geq 1$ may suffer from a case (ii) violation of the regularity condition.

4.1 Stationary, Stable VAR Models

4.1.1 Estimation[1]

Assume a stable VAR model of finite order p as defined in (2.3), (2.17) and (3.13):

$$\begin{aligned} y_t &= \Pi_1 y_{t-1} + \ldots + \Pi_p y_{t-p} + u_t \\ &= J\mathbf{\Pi}Y_{t-1} + JU_t \\ &= \Pi^{(1)} Y_{t-1} + u_t. \end{aligned}$$

Given T observations and p presample values, the model can be written in the form

$$y = \Pi^{(1)} Z + u, \tag{4.1}$$

with

$$\begin{aligned} y &:= [y_1, \ldots, y_T] && (k \times T), \\ \Pi^{(1)} &:= [\Pi_1, \ldots, \Pi_p] && (k \times kp), \\ Y_t &:= \begin{bmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix} && (kp \times 1), \\ Z &:= \left[Y_0, \ldots, Y_{T-1}\right] && (kp \times T), \\ u &:= [u_1, \ldots, u_T] && (k \times T). \end{aligned}$$

Least squares estimation of equation (4.1) yields

$$\hat{\Pi}^{(1)} = yZ'(ZZ')^{-1}. \tag{4.2}$$

Let the coefficients in $\hat{\Pi}^{(1)}$ be stacked into a $(k^2p \times 1)$ vector

$$\hat{\pi} = \text{vec}(\hat{\Pi}^{(1)}) = ((ZZ')^{-1}Z \otimes I_k)\text{vec}(y).$$

It is assumed throughout this chapter that the vector of error terms is a standard white noise process: $E(u_t) = 0$, $E(u_t u_t') = \Sigma_u$ a nonsingular, finite matrix, u_t and u_s are independent for $t \neq s$ and all fourth moments exist and are bounded (see Lütkepohl (1991, Definition 3.1)). Moreover, it is assumed that plim (ZZ'/T) converges to a

[1] See e.g. Lütkepohl (1991, Chapter 3), Davidson & MacKinnon (1993), Hamilton (1994, Chapter 11), Judge et al. (1988, Chapter 18.2), Johnston & DiNardo (1997, Chapter 9).

nonsingular, finite matrix. Under these assumptions, the LS estimator is consistent and asymptotically normally distributed as

$$\sqrt{T}(\hat{\pi} - \pi) \xrightarrow{d} N(0, \Sigma_{\hat{\pi}}) \quad \text{with} \quad \Sigma_{\hat{\pi}} = \text{plim}(ZZ'/T)^{-1} \otimes \Sigma_u. \tag{4.3}$$

The covariance matrix Σ_u can be consistently estimated as

$$\hat{\Sigma}_u = \frac{\hat{u}\hat{u}'}{T - kp} \quad \text{with} \quad \hat{u} = y - \hat{\Pi}^{(1)} Z,$$

and the covariance matrix $\Sigma_{\hat{\pi}}$ can be consistently estimated as

$$\hat{\Sigma}_{\hat{\pi}} = (ZZ'/T)^{-1} \otimes \hat{\Sigma}_u. \tag{4.4}$$

4.1.2 Standard Granger Noncausality

Let R be a $(p \times k^2 p)$ matrix which picks out of π all those p coefficients which capture the causal influence of $y_{1,t-i}$ on $y_{2,t}$ for $i = 1, \ldots, p$:

$$R = I_p \otimes [0,\, 1,\, 0,\, \ldots,\, 0]. \tag{4.5}$$

The null hypothesis that y_1 is Granger noncausal for y_2 at forecast horizon $h = 1$ can be written as

$$H_0^{\text{GC}} : \; g(\pi) = R\,\pi = 0, \tag{4.6}$$

and the corresponding Wald statistic is

$$\begin{aligned} \mathcal{W}_{\text{GC}} &= T g(\hat{\pi})' \, \hat{\Sigma}_{g(\hat{\pi})}^{-1} \, g(\hat{\pi}) \\ &= T(R\hat{\pi})' \left(R \hat{\Sigma}_{\hat{\pi}} R' \right)^{-1} (R\hat{\pi}). \end{aligned} \tag{4.7}$$

The Wald statistic has an asymptotic $\chi^2(p)$–distribution under the null hypothesis (4.6) if the $(p \times p)$ covariance matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi') \, \Sigma_{\hat{\pi}} \, (\partial g(\pi)'/\partial \pi)$ has full rank $r_g = p$. Since Z contains stationary regressors only and since $\text{plim}(ZZ'/T)$ converges to a nonsingular finite matrix by assumption, the asymptotic covariance matrix $\Sigma_{\hat{\pi}} = \text{plim}(ZZ'/T)^{-1} \otimes \Sigma_u$ is nonsingular. The rank of $\Sigma_{g(\hat{\pi})}$ is therefore equal to $\text{rk}(\partial g(\pi)/\partial \pi') = \text{rk}(R) = p$. Hence, the regularity condition is always fulfilled.
In small samples it is favorable to replace $\mathcal{W}_{\text{GC}}$ by $\widetilde{\mathcal{W}}_{\text{GC}} = \mathcal{W}_{\text{GC}}/p$. The latter statistic has an approximate $\mathcal{F}$–distribution with p and $(T - kp)$ degrees of freedom (Lütkepohl (1991, Chapter 3)).

4.1.3 Extended Granger Noncausality

Let $\pi^{(h)} = \text{vec}(\Pi^{(h)}) = \text{vec}(J\mathbf{\Pi}^h)$ with vec the column stacking operator and $J = [I_k, 0_k, \ldots, 0_k]$ a $(k \times kp)$ matrix. Moreover, define two k–dimensional vectors $R_1 = [1, 0, \ldots, 0]'$ and $R_2 = [0, 1, 0, \ldots, 0]'$, and let $R = I_p \otimes (R_1' \otimes R_2'))$. Then the null hypothesis that y_1 is Granger noncausal for y_2 at forecast horizon h can be written as

$$H_0^{\text{GCh}} : \; g(\pi) = R\,\pi^{(h)} = 0, \tag{4.8}$$

and the null hypothesis that y_1 is never Granger causal for y_2 is

$$H_0^{\text{EGC}} : \; g(\pi) = (I_{\tilde{h}} \otimes R)\,\boldsymbol{\pi}^{(\tilde{h})} = 0 \quad \text{with} \quad \boldsymbol{\pi}^{(\tilde{h})} = \begin{bmatrix} \pi^{(1)} \\ \vdots \\ \pi^{(\tilde{h})} \end{bmatrix}, \tag{4.9}$$

and $\tilde{h} = pk_3 + 1$.

Let $\hat{\Pi}^{(h)}$ be computed as $\Pi^{(h)}$ but with $\Pi_1, \ldots, \Pi_p$ replaced by their least squares estimates $\hat{\Pi}_1, \ldots, \hat{\Pi}_p$ (cf. (3.9) to (3.15), pages 32 to 33). Then $\hat{\pi}^{(h)} = \text{vec}(\hat{\Pi}^{(h)})$ is a consistent estimate of $\pi^{(h)}$.

Moreover, $\pi^{(h)}$ is a vector–valued continuously differentiable function with respect to π:

$$\begin{aligned} \frac{\partial \pi^{(h)}}{\partial \pi'} &= \frac{\partial \text{vec}(J\mathbf{\Pi}^h)}{\partial \pi'} \\ &= (I_{kp} \otimes J)\frac{\partial \text{vec}(\mathbf{\Pi}^h)}{\partial \pi'} \\ &= (I_{kp} \otimes J)\left[\sum_{i=0}^{h-1}(\mathbf{\Pi}')^{h-1-i} \otimes \mathbf{\Pi}^i\right]\frac{\partial \text{vec}(\mathbf{\Pi})}{\partial \pi'} \\ &= \sum_{i=0}^{h-1}(\mathbf{\Pi}')^{h-1-i} \otimes J\mathbf{\Pi}^i J' \end{aligned}$$

for $h > 1$. For $h = 1$, the matrix of partial derivatives $\partial\pi/\partial\pi' = I_{k^2p}$ (see Lütkepohl (1996a, Chapter 10), Lütkepohl & Burda (1997)). Hence, $\partial\pi^{(h)}/\partial\pi' \neq 0$ at π and it follows from Serfling (1980, p. 122-124), that

$$\sqrt{T}(\hat{\pi}^{(h)} - \pi^{(h)}) \quad \xrightarrow{d} \quad N(0, \Sigma_{\hat{\pi}}(h))$$

with

$$\Sigma_{\hat{\pi}}(h) = \frac{\partial \pi^{(h)}}{\partial \pi'}\Sigma_{\hat{\pi}}\frac{\partial \pi^{(h)'}}{\partial \pi}.$$

Let

$$\hat{\boldsymbol{\pi}}^{(\tilde{h})} = \begin{bmatrix} \hat{\pi} \\ \hat{\pi}^{(2)} \\ \vdots \\ \hat{\pi}^{(\tilde{h})} \end{bmatrix},$$

then the Wald statistic for a test of the null hypothesis (4.9) is

$$\begin{aligned} \mathcal{W}_{\text{EGC}}(\tilde{h}) &= T g(\hat{\pi})' \hat{\Sigma}_{g(\hat{\pi})}^{-1} g(\hat{\pi}) \\ &= T \left((I_{\tilde{h}} \otimes R) \hat{\boldsymbol{\pi}}^{(\tilde{h})} \right)' \left((I_{\tilde{h}} \otimes R) \frac{\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})}}{\partial \pi'} \hat{\Sigma}_{\hat{\pi}} \frac{\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})'}}{\partial \pi} (I_{\tilde{h}} \otimes R') \right)^{-1} \\ &\quad \times \left((I_{\tilde{h}} \otimes R) \hat{\boldsymbol{\pi}}^{(\tilde{h})} \right), \end{aligned} \tag{4.10}$$

with $\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})'} / \partial \pi$ computed as $\partial \boldsymbol{\pi}^{(\tilde{h})'} / \partial \pi = \left[I_{k^2 p}, \partial \pi^{(2)'} / \partial \pi, \ldots, \partial \pi^{(\tilde{h})'} / \partial \pi \right]$, but with the true VAR coefficients replaced by their least squares estimates.

Under the null hypothesis that y_1 is never Granger causal for y_2, the Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ has an asymptotic $\chi^2(\tilde{h}p)$–distribution if $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi') \Sigma_{\hat{\pi}} (\partial g(\pi)'/\partial \pi)$ is nonsingular.

Conditions under Which $\mathcal{W}_{\text{EGC}}(\tilde{h})$ May Not Have a Limiting χ^2–Distribution.

Since $\Sigma_{\hat{\pi}} = \text{plim}(ZZ'/T)^{-1} \otimes \Sigma_u$ is a nonsingular, finite matrix by assumption, $\Sigma_{g(\hat{\pi})}$ is regular unless the rank of the $(\tilde{h}p \times pk^2)$ matrix $(\partial g(\pi)/\partial \pi') = (I_{\tilde{h}} \otimes R)(\partial \boldsymbol{\pi}^{(\tilde{h})}/\partial \pi')$ is less than $\tilde{h}p$ under the null hypothesis.[2] If $\Sigma_{g(\hat{\pi})}$ is singular, the Wald statistic may no longer have a χ^2–distribution. However, the estimated covariance matrix $\hat{\Sigma}_{g(\hat{\pi})}$ may still have full row rank (see Andrews (1987)). In this case, computation of the Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ is feasible although a standard Wald test is no longer based on the usual limiting χ^2–distribution.

Example 4.1: Trivariate VAR(1) Model.

Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ holds under the null hypothesis (cf. Example 3.4)

$$H_0^{\text{EGC}} : \; g(\pi) = (I_2 \otimes R) \boldsymbol{\pi}^{(2)} = 0, \tag{4.11}$$

[2]Note, that for $\tilde{h}p > pk^2$, there are more restrictions than VAR coefficients and $\text{rk}(\partial g(\pi)/\partial \pi') < \tilde{h}p$ always holds, implying a singular covariance matrix $\Sigma_{g(\hat{\pi})}$. This case will not be considered here.

with

$$(I_2 \otimes R)\boldsymbol{\pi}^{(2)} = \begin{bmatrix} \pi_{21,1} \\ \pi^{(2)}_{21,1} \end{bmatrix} = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,1}\pi_{11,1} + \pi_{22,1}\pi_{21,1} + \pi_{23,1}\pi_{31,1} \end{bmatrix}.$$

The matrix of first order partial derivatives is

$$\frac{\partial g(\pi)}{\partial \pi'} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \pi_{21,1} & \pi_{11,1} + \pi_{22,1} & \pi_{23,1} & 0 & \pi_{21,1} & 0 & 0 & \pi_{31,1} & 0 \end{bmatrix}.$$

The null hypothesis (4.11) holds in the following cases:

$$\pi_{21,1} = 0 \quad \text{and} \quad \pi_{23,1} = 0, \ \pi_{31,1} \neq 0, \tag{4.12}$$

$$\pi_{21,1} = 0 \quad \text{and} \quad \pi_{23,1} \neq 0, \ \pi_{31,1} = 0, \tag{4.13}$$

$$\pi_{21,1} = 0 \quad \text{and} \quad \pi_{23,1} = 0, \ \pi_{31,1} = 0. \tag{4.14}$$

The Wald statistic has an asymptotic $\chi^2(\tilde{h}p)$–distribution if the regularity condition holds, i.e. if $\text{rk}(\partial g(\pi)/\partial \pi') = \tilde{h}p = 2$. It is easily seen that in case (4.14) $\text{rk}(\partial g(\pi)/\partial \pi') = 1 < \tilde{h}p$, and hence the rank condition is violated. The violation occurs because according to Corollary 3.4, each set of restrictions (4.12) or (4.13) is necessary and sufficient for y_1 being never Granger causal for y_2. Therefore, one restriction in (4.14) is superfluous. Including the redundant restriction into $g(\pi)$ leads to a singular covariance matrix $\Sigma_{g(\hat{\pi})}$. (See also Gallant (1977), Gallant & Tauchen (1989), Lütkepohl (1993) and Gaffke et al. (1999) for research on redundant restrictions.)

Example 4.1 illustrates that there exist parameter values under the null hypothesis which imply a singular covariance matrix $\Sigma_{g(\hat{\pi})}$. Proposition 4.1 states that for some VAR(p) models, $\Sigma_{g(\hat{\pi})}$ may always be singular under the null hypothesis.

Proposition 4.1:

For any stationary, stable VAR(p) model with $k_1 = k_2 = 1$, $k_3 \geq 1$ and $p \geq k$, the covariance matrix $\Sigma_{g(\hat{\pi})}$ is singular under H_0^{EGC} in (4.9).

Proof: Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ holds if

$$H_0^{\text{EGC}}: \; g(\pi) \;=\; (I_{\tilde{h}} \otimes R)\boldsymbol{\pi}^{(\tilde{h})} = \begin{bmatrix} \pi_{21,1} \\ \vdots \\ \pi_{21,p} \\ \pi_{21,1}^{(2)} \\ \vdots \\ \pi_{21,p}^{(2)} \\ \vdots \\ \pi_{21,p}^{(\tilde{h})} \end{bmatrix} = 0. \tag{4.15}$$

Using the recursion formula (3.17), page 34, the set of restrictions can be written as

$$\begin{aligned} g(\pi) \;&=\; G(\pi)\tilde{g}(\pi) \\ &= \begin{bmatrix} I_p & 0_p & \dots & \dots & 0_p \\ G_{21}(\pi) & I_p & 0_p & \dots & 0_p \\ G_{31}(\pi) & G_{32}(\pi) & I_p & \dots & 0_p \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ G_{\tilde{h}1}(\pi) & G_{\tilde{h}2}(\pi) & \dots & \dots & I_p \end{bmatrix} \begin{bmatrix} \pi_{21,1} \\ \vdots \\ \pi_{21,p} \\ \pi_{23,1}\pi_{31,1} \\ \vdots \\ \pi_{23,1}\pi_{31,p} \\ \pi_{23,1}^{(2)}\pi_{31,1} \\ \vdots \\ \pi_{23,1}^{(\tilde{h}-1)}\pi_{31,p} \end{bmatrix}, \end{aligned}$$

with $G(\pi)$ a $(\tilde{h}p \times \tilde{h}p)$ lower triangular matrix where 0_p denotes a $(p \times p)$ matrix of zeros and $G_{lm}(\pi)$ denotes a $(p \times p)$ submatrix whose elements are possibly nonlinear functions of the elements of π, and $\tilde{g}(\pi)$ a $(\tilde{h}p \times 1)$ vector. An illustration of the reformulation is given in Appendix A.

The matrix of first order partial derivatives is (Lütkepohl (1996a, 10.5.5 (3))):

$$\frac{\partial g(\pi)}{\partial \pi'} \;=\; G(\pi)\frac{\partial \tilde{g}(\pi)}{\partial \pi'} + \left(\tilde{g}(\pi)' \otimes I_{\tilde{h}p}\right)\frac{\partial \text{vec}(G(\pi))}{\partial \pi'}. \tag{4.16}$$

Under the null hypothesis (4.15), $\tilde{g}(\pi) = 0$ (see Corollary 3.3). Moreover, $G(\pi)$ is a nonsingular matrix. Therefore, $\text{rk}(\partial g(\pi)/\partial \pi') = \text{rk}(\partial \tilde{g}(\pi)/\partial \pi')$ under the null hypothesis (see Lütkepohl (1996a, 4.3.1.(9))).

To show that $\text{rk}(\partial\tilde{g}(\pi)/\partial\pi') < \tilde{h}p$ under the null hypothesis H_0^{EGC} in (4.9), it is convenient to write

$$\begin{aligned}\tilde{g}(\pi) &= F(\pi)f(\pi) &(4.17)\\ &= \begin{bmatrix} I_p & 0_{p\times k_3 p} \\ 0_p & I_p \otimes \pi_{23,1} \\ 0_p & I_p \otimes \pi^{(2)}_{23,1} \\ \vdots & \vdots \\ 0_p & I_p \otimes \pi^{(\tilde{h}-1)}_{23,1} \end{bmatrix} \begin{bmatrix} \pi_{21,1} \\ \vdots \\ \pi_{21,p} \\ \pi_{31,1} \\ \vdots \\ \pi_{31,p} \end{bmatrix}, &(4.18)\end{aligned}$$

with $0_{p\times k_3 p}$ a $(p \times k_3 p)$ matrix of zeros, $F(\pi)$ a $(\tilde{h}p \times (k_3+1)p)$ matrix of rank $(k_3+1)p$ and $f(\pi)$ a $((k_3+1)p \times 1)$ vector. With this notation, the matrix of first order partial derivatives can be written as:

$$\frac{\partial\tilde{g}(\pi)}{\partial\pi'} = F(\pi)\frac{\partial f(\pi)}{\partial\pi'} + \left(f(\pi)' \otimes I_{\tilde{h}p}\right)\frac{\partial \text{vec}(F(\pi))}{\partial\pi'}. \tag{4.19}$$

The following inequality gives an upper bound for the rank of the matrix of first order partial derivatives (see Lütkepohl (1996a, 4.3.3 (2a), (3))):

$$\text{rk}\left(\frac{\partial\tilde{g}(\pi)}{\partial\pi'}\right) \leq \text{rk}\left(F(\pi)\frac{\partial f(\pi)}{\partial\pi'}\right) + \text{rk}\left(\left(f(\pi)' \otimes I_{\tilde{h}p}\right)\frac{\partial \text{vec}(F(\pi))}{\partial\pi'}\right). \tag{4.20}$$

Now

$$\begin{aligned}\text{rk}\left(F(\pi)\frac{\partial f(\pi)}{\partial\pi'}\right) &\leq \min\left\{\text{rk}\left(F(\pi)\right), \text{rk}\left(\frac{\partial f(\pi)}{\partial\pi'}\right)\right\}\\ &\leq (k_3+1)p,\end{aligned}$$

since $\partial f(\pi)/\partial\pi'$ is a $((k_3+1)p \times pk^2)$ matrix with $\text{rk}(\partial f(\pi)/\partial\pi') \leq (k_3+1)p$ and $\text{rk}(F(\pi)) \leq (k_3+1)p$. Moreover,

$$\begin{aligned}\text{rk}\left(\left(f(\pi)' \otimes I_{\tilde{h}p}\right)\frac{\partial \text{vec}(F(\pi))}{\partial\pi'}\right) &= \min\left\{\text{rk}\left(f(\pi)' \otimes I_{\tilde{h}p}\right), \text{rk}\left(\frac{\partial \text{vec}(F(\pi))}{\partial\pi'}\right)\right\}\\ &\leq (\tilde{h}-1)k_3,\end{aligned}$$

since $\text{vec}(F(\pi))$ contains only $(\tilde{h}-1)k_3$ different elements captured in $\pi^{(h)}_{23,1}$, $h = 1,\ldots,\tilde{h}-1$, which are a function of the parameters in π. Inserting into (4.20) yields

$$\text{rk}\left(\frac{\partial\tilde{g}(\pi)}{\partial\pi'}\right) \leq (k_3+1)p + (\tilde{h}-1)k_3 = p + k_3 p(k_3+1). \tag{4.21}$$

Note, that $\tilde{g}(\pi)$ contains p coefficients $\pi_{21,i},\ i = 1, \ldots, p$ and $(\tilde{h}-1)k_3 = pk_3^2$ coefficients $\pi_{23,1}^{(h)},\ h = 1, \ldots, \tilde{h}-1$ as well as pk_3 coefficients $\pi_{31,i},\ i = 1, \ldots, p$. Hence, altogether $\tilde{g}(\pi)$ contains $p+pk_3(k_3+1)$ different coefficients. The inequality (4.21) therefore states that the rank of the matrix of partial derivatives is less than or equal to the number of different elements of $\tilde{g}(\pi)$. It follows that $\text{rk}(\partial\tilde{g}(\pi)/\partial\pi') < \tilde{h}p$ whenever

$$\begin{aligned} & p + pk_3(k_3+1) < \tilde{h}p \\ \leftrightarrow\ & 1 + k_3(k_3+1) < pk_3 + 1 \\ \leftrightarrow\ & k_3 + 1 < p. \end{aligned}$$

Hence, for any VAR(p) model with $k_1 = k_2 = 1$ and $p \geq k$,

$$\text{rk}(\partial g(\pi)/\partial\pi') \ = \ \text{rk}(\partial\tilde{g}(\pi)/\partial\pi') < \tilde{h}p, \tag{4.22}$$

under the null hypothesis H_0^{EGC} such that $\Sigma_{g(\hat{\pi})}$ is singular. □

Note, that Proposition 4.1 only gives an upper bound for p: a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ may arise under the null hypothesis (4.9) even if $p < k$. For instance, for $k_1 = k_2 = k_3 = 1$, Proposition 4.1 states that a singularity occurs whenever $p \geq 3$. However, Example 4.1 has already shown that a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ may also arise in a VAR(1) model under some parameter values. Moreover, Corollary 4.1 shows that in trivariate VAR models, $\Sigma_{g(\hat{\pi})}$ is always singular under the null hypothesis (4.9) if $p \geq 2$.

Corollary 4.1:
For any stationary, stable VAR(p) model with $k_1 = k_2 = k_3 = 1$ and $p \geq 2$, the covariance matrix $\Sigma_{g(\hat{\pi})}$ is singular under the null hypothesis (4.9).

Proof: Two sets of necessary and sufficient restrictions for Granger noncausality of y_1 for y_2 at all forecast horizons are given in Corollary 3.3:
If (3.19) holds, then $\text{rk}(F(\pi)) = p$ and

$$\begin{aligned} \text{rk}\left(F(\pi)\frac{\partial f(\pi)}{\partial\pi'}\right) \ &\leq\ \min\left\{\text{rk}\left(F(\pi)\right), \text{rk}\left(\frac{\partial f(\pi)}{\partial\pi'}\right)\right\} \\ &=\ p. \end{aligned}$$

Moreover, since $k_3 = 1$, $\tilde{h} = p + 1$ and

$$\begin{aligned} \operatorname{rk}\left(\left(f(\pi)' \otimes I_{\tilde{h}p}\right) \frac{\partial \operatorname{vec}(F(\pi))}{\partial \pi'}\right) &\leq \min\left\{\operatorname{rk}\left(f(\pi)' \otimes I_{\tilde{h}p}\right), \operatorname{rk}\left(\frac{\partial \operatorname{vec}(F(\pi))}{\partial \pi'}\right)\right\} \\ &\leq (\tilde{h}-1)k_3 \\ &\leq p. \end{aligned}$$

Inserting into (4.20) yields

$$\operatorname{rk}\left(\frac{\partial \tilde{g}(\pi)}{\partial \pi'}\right) \leq 2p.$$

Alternatively, if (3.20) holds, $f(\pi) = 0$ under the null hypothesis. Inserting into (4.20) yields

$$\begin{aligned} \operatorname{rk}\left(\frac{\partial \tilde{g}(\pi)}{\partial \pi'}\right) &\leq \operatorname{rk}\left(F(\pi)\frac{\partial f(\pi)}{\partial \pi'}\right) \\ &\leq \min\left\{\operatorname{rk}\left(F(\pi)\right), \operatorname{rk}\left(\frac{\partial f(\pi)}{\partial \pi'}\right)\right\} \\ &\leq (1 + k_3)p \\ &\leq 2p, \end{aligned}$$

since $k_3 = 1$ for the present example. Hence, $\operatorname{rk}(\partial \tilde{g}(\pi)/\partial \pi') < \tilde{h}p$ whenever

$$\begin{aligned} & 2p < \tilde{h}p \\ \leftrightarrow\ & 2 < pk_3 + 1 \\ \leftrightarrow\ & 1 < p. \end{aligned}$$

It follows that in stationary VAR(p) models with $k_1 = k_2 = k_3 = 1$ and $p > 1$, $\Sigma_{g(\hat{\pi})}$ is singular under the null hypothesis that y_1 is never Granger causal for y_2. □

Causes for a Singular Covariance Matrix $\Sigma_{g(\hat{\pi})}$ under H_0^{EGC}.

Proposition 4.1 and Corollary 4.1 show that a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ under the null hypothesis H_0^{EGC} is rather the rule than the exception, at least in trivariate VAR models with $k_1 = k_2 = k_3 = 1$. Singularity arises for three reasons:

First, for $k_3 \geq 1$, there are 2^{k_3} possibilities how indirect causality may be transmitted from y_1 to y_2 via causal chains inside y_3. (In other words, there exist 2^{k_3} possibilities to split up the k variables in y_t among two vectors $S_{1,t}$, $S_{2,t}$ as illustrated in Example

3.2, page 29.) The extended Granger causality concept translates these 2^{k_3} possibilities into $(\tilde{h}-1)p$ sets of nonlinear restrictions on the $k_3(\tilde{h}-1)$ parameters in $\pi_{23,1}^{(h)}$, $h = 1, \ldots, \tilde{h}-1$, and on the $k_3 p$ parameters in $\pi_{31,i}$, $i = 1, \ldots, p$. However, for $p \geq k$, $(\tilde{h}-1)p = p^2 k_3 > pk_3(k_3+1)$, hence there are more restrictions than parameters involved. As consequence, some restrictions are redundant under the null hypothesis. Second, the maximal forecast horizon $\tilde{h} = pk_3 + 1$ is an upper bound at which an indirect Granger causal link will be detected at the latest. However, for some parameter values under the null hypothesis, Granger noncausality up to a smaller forecast horizon $h < \tilde{h}$ already implies that y_1 is never causal for y_2. Hence, under these parameter values $g(\pi)$ contains redundant restrictions, see Dufour & Renault (1998, p. 1115).[3] Third, due to the nonlinear nature of the restrictions, superfluous restrictions may arise for some parameter values under the null hypothesis as illustrated in Example 4.1, equation (4.14). Note, that this problem also applies to a standard Wald test that all responses of y_2 to an impulse in y_1 are zero. In contrast, the first and second argument hold only for a standard Wald test that y_1 is never causal for y_2, hence if the whole set of $\tilde{h}p$ noncausality restrictions in (4.9) is tested.

A possibly singular covariance matrix $\Sigma_{g(\hat{\pi})}$ may also occur under the null hypothesis H_0^{GCh} (see (4.8) on page 65) that y_1 is Granger noncausal for y_2 *at* forecast horizon $h > 1$. To see this, consider again Example 4.1: Granger noncausality of y_1 for y_2 at forecast horizons $h = 1, 2$ is characterized by the two restrictions

$$H_0^{\mathrm{EGC}}: \; g(\pi) = \begin{bmatrix} g_1(\pi) \\ g_2(\pi) \end{bmatrix} = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,1}^{(2)} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Example 4.1 has shown that if the restrictions in $g(\pi)$ are tested jointly, the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has reduced row rank equal to one under H_0^{EGC} if the true data generating process is (4.14). If on the other hand, the restrictions in $g(\pi)$ are tested separately, the Jacobian matrix $\partial g_1(\pi)/\partial \pi'$ will always have full row rank. However, the rank of the Jacobian matrix

$$\partial g_2(\pi)/\partial \pi' = \begin{bmatrix} \pi_{21,1} & \pi_{11,1}+\pi_{22,1} & \pi_{23,1} & 0 & \pi_{21,1} & 0 & 0 & \pi_{31,1} & 0 \end{bmatrix}$$

[3] Redundant restrictions occur for instance in Example 3.4 under conditions (i) and (ii). Note, however, that under either condition $\partial g(\pi)/\partial \pi'$ has full row rank unless in addition either $\pi_{31,1} = 0$ or $\pi_{41,1} = 0$ holds (with condition (i)) or $\pi_{23,1} = 0$ or $\pi_{24,1} = 0$ holds (with condition (ii)).

may be a zero vector under H_0^{EGC}, if (4.14) holds and $\pi_{11,1} + \pi_{22,1} = 0$.

Alternative Formulations of the Null Hypothesis.

Since the Wald statistic is not invariant to algebraically equivalent reformulations of the null hypothesis (see e.g. Gregory & Veall (1985), Lafontaine & White (1986), Breusch & Schmidt (1988), Critchley et al. (1996)), one might ask for alternative reformulations of the restrictions for Granger noncausality at all forecast horizons. However, even for the simple trivariate VAR(1) model considered in Example 4.1, no reformulation seems to be at hand which avoids nonlinear restrictions.[4]

Linear restrictions arise, however, if the set of k variables in y_t is split up into two vectors $S_1 = [y_1, \bar{y}_3']'$ and $S_2 = [y_2, \underline{y}_3']'$ where $y_3 = \bar{y}_3 \cup \underline{y}_3$ and $\bar{y}_3 \cap \underline{y}_3 = \{\}$. If S_1 is Granger noncausal for S_2 at forecast horizon $h = 1$, then all variables in S_1 are never Granger causal for any variables in S_2. However, there exist 2^{k_3} different pairs of vectors S_1, S_2. Checking all possible combinations leads to a sequential test of nonnested hypotheses. The idea is briefly discussed in the context of Chapter 5 (see Figure 5.2 for a sequential testing strategy for a trivariate VAR(p) model).

Problems Related to Likelihood Ratio and Lagrange Multiplier Tests.

Using Likelihood Ratio or Lagrange Multiplier tests instead of the standard Wald test is not a way out: if the Jacobian matrix of the set of restrictions $\partial g(\pi)/\partial \pi'$ has reduced rank under the null hypothesis, the restricted (pseudo) maximum likelihood estimator cannot be determined. This will be demonstrated for the case of the Lagrange Multiplier test statistic (see e.g. Judge et al. (1988, Chapter 3) and Davidson & MacKinnon (1993, Chapters 8 and 13)):

Applying the vec operator to equation (4.1), page 63, yields

$$\text{vec}(y) = (Z' \otimes I_k)\pi + \text{vec}(u), \tag{4.23}$$

where it is assumed that $\text{vec}(u) \dot{\sim} N(0, (I_T \otimes \Sigma_u))$ is asymptotically normally distributed. Apart from a constant, the log likelihood function can be written as

$$\ln l(\pi, \Sigma_u) = -\frac{T}{2}\ln|\Sigma_u| \tag{4.24}$$

[4]See Bruneau & Nicolai (1992a) and Dufour & Renault (1998) for alternative sets of restrictions for Granger noncausality at all forecast horizons.

$$- \quad \frac{1}{2}\left(\text{vec}(y) - (Z' \otimes I_k)\pi\right)' \left(I_T \otimes \Sigma_u^{-1}\right) \left(\text{vec}(y) - (Z' \otimes I_k)\pi\right).$$

Let $\tilde{\pi}_0$ denote the restricted (pseudo) maximum likelihood estimator which fulfills the restrictions under H_0^{EGC}, i.e. $g(\tilde{\pi}_0) = 0$. The restricted (pseudo) ML estimator is found by maximizing the Lagrangian function

$$h(\pi, \Sigma_u, \eta) \quad = \quad \ln \text{l}(\pi, \Sigma_u) + g(\pi)'\eta,$$

with η a $(\tilde{h}p \times 1)$ vector of Lagrange multipliers. The first order conditions for a maximum are

$$\frac{\partial h(\pi, \Sigma_u, \eta)}{\partial \pi} \quad = \quad \frac{\partial \ln \text{l}(\pi, \Sigma_u)}{\partial \pi} + \frac{\partial g(\pi)'}{\partial \pi}\eta \stackrel{!}{=} 0, \tag{4.25}$$

and

$$g(\pi) \quad \stackrel{!}{=} \quad 0. \tag{4.26}$$

Condition (4.25) can be written as

$$(Z \otimes \Sigma_u^{-1})\text{vec}(y) - (ZZ' \otimes \Sigma_u^{-1})\tilde{\pi}_0 + \tilde{S}_0\eta_0 = 0, \tag{4.27}$$

where

$$\tilde{S}_0 \quad = \quad \frac{\partial g(\pi)'}{\partial \pi}|_{\pi=\tilde{\pi}_0}.$$

Left–multiplying (4.27) with $\tilde{S}_1 = ((ZZ')^{-1} \otimes \Sigma_u)$ yields

$$\tilde{\pi}_0 = \tilde{\pi} + \tilde{S}_1\tilde{S}_0\eta_0, \tag{4.28}$$

with $\tilde{\pi}$ the unrestricted (pseudo) maximum likelihood estimator (which is asymptotically identical to the least squares estimator). Left–multiplying (4.28) with $\tilde{S}_0'$ and solving for η_0 yields

$$\eta_0 = (\tilde{S}_0'\tilde{S}_1\tilde{S}_0)^{-1}\tilde{S}_0'(\tilde{\pi}_0 - \tilde{\pi}), \tag{4.29}$$

iff the $(\tilde{h}p \times \tilde{h}p)$ matrix $(\tilde{S}_0'\tilde{S}_1\tilde{S}_0)$ is nonsingular such that the inverse exists.

The restricted maximum likelihood estimator $\tilde{\pi}_0$ is then found by inserting (4.29) into (4.28):

$$\tilde{\pi}_0 = \tilde{\pi} + \tilde{S}_1\tilde{S}_0(\tilde{S}_0'\tilde{S}_1\tilde{S}_0)^{-1}\tilde{S}_0'\tilde{\pi}. \tag{4.30}$$

However, if the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has reduced rank under the null hypothesis, $(\tilde{S}_0'\tilde{S}_1\tilde{S}_0)$, evaluated at the true parameters under H_0^{EGC}, is not invertible in (4.29) and the restricted ML estimator cannot be uniquely determined. This problem also holds for the Likelihood Ratio test statistic. (See also Lütkepohl (1993)).

4.1.4 Impulse Response Analysis

According to Proposition 3.3, y_1 is never causal for y_2 in terms of impulse response analysis if

$$H_0^{\text{IR}}: \ \phi_{21,j} = 0 \quad \forall\, j = 1, \ldots, \bar{h}, \tag{4.31}$$

with $\bar{h} = p(k-1)$. The impulse response coefficients $\phi_{21,j}$ can be expressed as nonlinear functions of the vector autoregressive coefficients in π: let

$$\phi_j \ = \ \text{vec}(\Phi_j) = \text{vec}(\Pi_1^{(j)}) = (J \otimes I_k)\pi^{(j)},$$

with $J = [I_k, 0, \ldots, 0]$ a $(k \times kp)$ matrix. Define two k–dimensional vectors $R_1 = [1, 0, \ldots, 0]'$ and $R_2 = [0, 1, 0, \ldots, 0]'$, then

$$\phi_{21,j} \ = \ (R_1' \otimes R_2')\phi_j = (R_1' J \otimes R_2')\pi^{(j)}.$$

Let

$$\boldsymbol{\phi}^{(\bar{h})} \ = \ \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_{\bar{h}} \end{bmatrix},$$

then the null hypothesis (4.31) can be written in terms of the vector autoregressive coefficients as

$$H_0^{\text{IR}}: g(\pi) \ = \ (I_{\bar{h}} \otimes R_1' J \otimes R_2')\boldsymbol{\pi}^{(\bar{h})} = (I_{\bar{h}} \otimes \bar{R})\boldsymbol{\pi}^{(\bar{h})}, \tag{4.32}$$

where $\bar{R} = (R_1' J \otimes R_2')$.

The Wald statistic for a test of the null hypothesis (4.32) can be set up as in (4.11) on page 66 but with R replaced by $\bar{R}$ and $\tilde{h}$ replaced by $\bar{h}$:

$$\begin{aligned} \mathcal{W}_{\text{IR}}(\bar{h}) \ &= \ T g(\hat{\pi})' \, \hat{\Sigma}_{g(\hat{\pi})}^{-1} \, g(\hat{\pi}) \\ &= \ T \left((I_{\bar{h}} \otimes \bar{R})\hat{\boldsymbol{\pi}}^{(\bar{h})} \right)' \left((I_{\bar{h}} \otimes \bar{R}) \frac{\partial \hat{\boldsymbol{\pi}}^{(\bar{h})}}{\partial \pi'} \hat{\Sigma}_{\hat{\pi}} \frac{\partial \hat{\boldsymbol{\pi}}^{(\bar{h})'}}{\partial \pi} (I_{\bar{h}} \otimes \bar{R}') \right)^{-1} \\ &\quad \times \left((I_{\bar{h}} \otimes \bar{R})\hat{\boldsymbol{\pi}}^{(\bar{h})} \right). \end{aligned} \tag{4.33}$$

Under the null hypothesis (4.32), the Wald statistic $\mathcal{W}_{\text{IR}}(\bar{h})$ has an asymptotic $\chi^2(\bar{h})$–distribution if $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi')\Sigma_{\hat{\pi}}(\partial g(\pi)'/\partial \pi)$ is nonsingular.

Causes for a Singular Covariance Matrix $\Sigma_{g(\hat{\pi})}$ under $H_0^{\mathbf{IR}}$.

The discussion of the standard Wald test of Granger noncausality at all forecast horizons $h \geq 1$ has already shown that nonlinear restrictions may cause a reduced row rank of the Jacobian matrix $\partial g(\pi)/\partial \pi'$ under some parameter values under H_0 and hence a singular covariance matrix $\Sigma_{g(\hat{\pi})}$. Let us therefore take a closer look at the restrictions which characterize noncausality of y_1 for y_2 in terms of impulse response analysis:

$$H_0^{\text{IR}}: \; g(\pi) = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,1}^{(2)} \\ \vdots \\ \pi_{21,1}^{(\bar{h})} \end{bmatrix} = 0. \tag{4.34}$$

Inserting restrictions succinctly, using the recursion formula (3.17) on page 34, yields

$$H_0^{\text{IR}}: \; g(\pi) = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,2} + \pi_{23,1}\pi_{31,1} \\ \pi_{21,2}^{(2)} + \pi_{23,1}^{(2)}\pi_{31,1} \\ \vdots \\ \pi_{21,2}^{(\bar{h}-1)} + \pi_{23,1}^{(\bar{h}-1)}\pi_{31,1} \end{bmatrix} = 0. \tag{4.35}$$

For VAR(p) models with $p > 1$, the restrictions at forecast horizons $h > 1$ consist of two additive terms, $\pi_{21,2}^{(h-1)}$ and $(\pi_{23,1}^{(h-1)}\pi_{31,1})$. The latter term is always a nonlinear function of the coefficients in π. The former term is linear for $h = 2$; for $h \leq p$, it can be written as the sum of a linear term $(\pi_{21,h})$ and a nonlinear function of the coefficients in π. For $h > p$, the set of restrictions are a purely nonlinear function of the coefficients in π.

Since the restrictions consist of nonlinear functions of the VAR coefficients, the rank of the Jacobian matrix $\partial g(\pi)/\partial \pi'$ may be less than $\bar{h}$ for some parameter values under H_0^{IR} (see Lütkepohl (1993) and Lütkepohl & Breitung (1997)). On the other hand, the restrictions in (4.35) include an additive linear term up to forecast horizon $h = p$. Therefore, $p \leq \text{rk}(\partial g(\pi)/\partial \pi') \leq \bar{h}$.

An illustration is given in Example 4.1 which carries over to impulse response analysis: if $p = 1$, $\tilde{h}p = \tilde{h} = \bar{h}$ and $R = \bar{R}$, so that zero responses of y_2 to a one–time, one–unit shock in y_1 and Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$ are both characterized by the same set of restrictions (see Lemma 3.3 and Lütkepohl (1982,

1993)).

Note, that Proposition 4.1 and Corollary 4.1 do not carry over to impulse response analysis. Both are based on the condition that

$$\begin{aligned} \pi_{21,i}^{(h)} &= 0 \quad \forall\ h = 1, \ldots, \tilde{h} \quad \leftrightarrow \\ \pi_{21,i} &= 0 \text{ and } \pi_{23,1}^{(h-1)} \pi_{31,i} = 0 \quad \forall\ i = 1, \ldots, p \text{ and } h = 2, \ldots, \tilde{h}. \end{aligned}$$

However, this condition need not be fulfilled under H_0^{IR}.

Alternative Formulations of the Null Hypothesis.

An alternative to testing H_0^{IR} may be seen in testing whether the $\bar{h}$–th interim multiplier equals zero, i.e.

$$\begin{aligned} H_0^{\text{IM}} :\ \phi_{21}(\bar{h}) &= \sum_{j=1}^{\bar{h}} \phi_{21,j} \\ &= \sum_{j=1}^{\bar{h}} \pi_{21,1}^{(j)}. \end{aligned}$$

Since the restriction under H_0^{IM} contains one additive linear term ($\pi_{21,1}$), the Jacobian matrix has always full rank equal to one under the null hypothesis. Moreover, H_0^{IR} implies H_0^{IM} but not vice versa. Hence, if a standard Wald test rejects the latter null hypothesis, it implicitly rejects the former null hypothesis. However, if H_0^{IM} cannot be rejected, no conclusion can be drawn for a standard Wald test of H_0^{IR} (cf. page 40).

4.2 Nonstationary VAR Models

4.2.1 Estimation[5]

If the vector of variables $y_t \sim I(1)$ and cointegrated with cointegration rank r, $0 < r < k$, the levels VAR model can be written in error correction form (2.6), page 9, or as

$$Z_{0t} = AB'Z_{1t} + \Psi Z_{2t} + u_t, \tag{4.36}$$

with

$$\begin{aligned} Z_{0t} &:= \Delta y_t, \\ Z_{1t} &:= y_{t-1}, \\ Z_{2t} &:= \begin{bmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{bmatrix}, \\ \Pi(1) &:= AB', \\ \Psi &:= \left[\Gamma_1, \ldots, \Gamma_{p-1}\right]. \end{aligned}$$

Under the assumption that $u_t \sim N(0, \Sigma_u)$ with Σ_u a nonsingular covariance matrix, (4.36) can be estimated by Johansen's ML procedure (see Johansen (1988) and Johansen (1991)). Thereby, the maximum likelihood estimates $\tilde{A}$, $\tilde{B}$, $\tilde{\Psi}$ and $\tilde{\Sigma}_u$ are found by maximizing the log likelihood function

$$\begin{aligned} \mathrm{l}(A, B, \Psi, \Sigma_u) &= -\frac{T}{2}\ln|\Sigma_u| \\ &- \frac{1}{2}\sum_{t=1}^{T}(Z_{0t} - AB'Z_{1t} - \Psi Z_{2t})'\Sigma_u^{-1}(Z_{0t} - AB'Z_{1t} - \Psi Z_{2t}), \end{aligned}$$

subject to the rank restriction $\mathrm{rk}(\Pi(1)) = \mathrm{rk}(AB') = r$.

Let

$$\begin{aligned} R_{it} &= Z_{it} - \sum_{t=1}^{T} Z_{it}Z_{2t}' \left[\sum_{t=1}^{T}(1/T)Z_{2t}Z_{2t}'\right]^{-1}, \\ S_{ij} &= \frac{1}{T}\sum_{t=1}^{T} R_{it}R_{jt}', \end{aligned}$$

[5] See e.g. Lütkepohl (1991, 1999), Banerjee et al. (1993), Johansen (1995) and Hatanaka (1996).

for $i, j = 0, 1$. Then maximizing the log likelihood function can be translated into solving

$$|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}| = 0 \tag{4.37}$$

for eigenvalues $1 > \hat{\lambda}_1 > \ldots > \hat{\lambda}_k > 0$ and into solving

$$\left[\lambda_i S_{11} - S_{10}S_{00}^{-1}S_{01}\right] v_i = 0 \tag{4.38}$$

for the corresponding eigenvectors $\hat{V} = [\hat{v}_1, \ldots, \hat{v}_k]$, subject to the restriction $\hat{V}' S_{11} \hat{V} = I_k$. This yields the maximum likelihood estimate

$$\tilde{B} = [\tilde{v}_1, \ldots, \tilde{v}_r]$$

see Theorem 6.1 of Johansen (1995). Given $\tilde{B}$,

$$\begin{aligned}
\tilde{A} &= S_{01}\tilde{B}(\tilde{B}' S_{11} \tilde{B})^{-1}, \\
\tilde{\Psi} &= (Z_{0t} - R_{0t}) - \tilde{A}\tilde{B}(Z_{1t} - R_{1t}), \\
\tilde{\Sigma}_u &= (1/T)\sum_{t=1}^{T}(Z_{0t} - \tilde{A}\tilde{B}Z_{1t} - \tilde{\Psi}Z_{2t})(Z_{0t} - \tilde{A}\tilde{B}Z_{1t} - \tilde{\Psi}Z_{2t})'.
\end{aligned}$$

Under the assumptions that u_t is a Gaussian white noise process, that $y_t \sim I(1)$ and cointegrated with known cointegration rank r, $0 < r < k$ and that the initial values $y_0, \ldots, y_{-p}$ are fixed, the maximum likelihood estimators $\tilde{\Pi}(1) = \tilde{A}\tilde{B}$, $\tilde{\Psi}$ and $\tilde{\Sigma}_u$ are consistent estimators.

Let $\Psi^\star = [\Psi, \Pi(1)]$, $\psi = \mathrm{vec}(\Psi)$ and $\psi^\star = \mathrm{vec}(\Psi^\star)$. Moreover, let $\tilde{\Psi}^\star$, $\tilde{\psi}$ and $\tilde{\psi}^\star$ denote the corresponding ML estimators, then

$$\sqrt{T}(\tilde{\psi}^\star - \psi^\star) \xrightarrow{d} N(0, \Sigma_{\tilde{\psi}^\star}), \tag{4.39}$$

with $\Sigma_{\tilde{\psi}^\star}$ a singular covariance matrix. However,

$$\sqrt{T}(\tilde{\psi} - \psi) \xrightarrow{d} N(0, \Sigma_{\tilde{\psi}}), \tag{4.40}$$

where $\Sigma_{\tilde{\psi}}$ is the upper left–hand, nonsingular $(k^2(p-1) \times k^2(p-1))$ submatrix of $\Sigma_{\tilde{\psi}^\star}$ in (4.39). The singularity of $\Sigma_{\tilde{\psi}^\star}$ arises because some coefficients in $\Pi(1)$ cannot be written as coefficients attached to stationary regressors. As consequence, the estimators of these coefficients converge at a rate faster than $T^{1/2}$, leading to a singular covariance matrix.

Often hypotheses are formulated with respect to the coefficients in the loading matrix A or the cointegrating matrix B alone. Note, that the decomposition $\Pi(1) = AB'$ is not unique unless the $(r \times k)$ matrices A, B have been normalized in some way. For the same reason, the estimates $\tilde{A}, \tilde{B}$ are not unique and consistency does not hold without further identifying restrictions (see Lütkepohl (1991, Chapter 11)). Under uniqueness restrictions, $\sqrt{T}\text{vec}(\tilde{A} - A)$ converges to a normal distribution while $T\text{vec}(\tilde{B} - B)$ converges to a mixed normal distribution (see Johansen (1995, Chapter 13)).

Estimates of the vector autoregressive matrices $\Pi_1, \ldots, \Pi_p$ can be computed from the ML estimates of $\tilde{\Pi}(1)$ and $\tilde{\Psi} = [\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{p-1}]$ as

$$\begin{aligned}
\tilde{\Pi}_1 &= -\tilde{\Pi}(1) + I_k + \tilde{\Gamma}_1, \\
\tilde{\Pi}_j &= \tilde{\Gamma}_j - \tilde{\Gamma}_{j-1}, \quad j = 2, \ldots, p-1, \\
\tilde{\Pi}_p &= -\tilde{\Gamma}_{p-1}.
\end{aligned}$$

Since the matrices $\tilde{\Pi}_i,\ i = 1, \ldots, p$, can be written as a linear combination of the matrix $\tilde{\Psi}^\star = [\tilde{\Psi}, \tilde{\Pi}(1)]$, and since a linear combination of normally distributed variables is also normally distributed, it follows from (4.39) that

$$\sqrt{T}(\tilde{\pi} - \pi) \xrightarrow{d} N(0, \Sigma_{\tilde{\pi}}) \tag{4.41}$$

with $\tilde{\pi} = \text{vec}(\tilde{\Pi}^{(1)})$ and $\tilde{\Pi}^{(1)} = [\tilde{\Pi}_1, \ldots, \tilde{\Pi}_p]$. For $0 < r < k$, the covariance matrix $\Sigma_{\tilde{\pi}}$ is singular.

Testing for the Cointegration Rank.

The Johansen ML procedure outlined above yields maximum likelihood estimates conditional on a known cointegration rank r. In practice, r is unknown but can be determined by a likelihood ratio test (see Johansen (1995, Chapter 6), Lütkepohl (1991, Chapter 11), Banerjee et al. (1993, Chapter 8)). A test of the hypothesis $H_0 : r = r_0$ against $H_1 : r_0 < r \leq r_1$ is based on the statistic

$$\text{LR}(r_0, r_1) = 2\left(\text{l}(r_1) - \text{l}(r_0)\right) = -T \sum_{i=r_0+1}^{r_1} \ln(1 - \tilde{\lambda}_i),$$

where $\text{l}(r_0),\ \text{l}(r_1)$ denote the value of the log likelihood function obtained under the null and the alternative hypothesis and $\tilde{\lambda}_i$ is determined by equation (4.37). Setting $r_1 = k$

yields the trace test while choosing $r_1 = r_0 + 1$ results in the maximal eigenvalue test. The asymptotic distribution of the test statistic $\mathrm{LR}(r_0, r_1)$ is nonstandard and depends on the difference $(r_1 - r_0)$ and on deterministic terms included in the data generating process. Critical values have been tabulated e.g. by Johansen (1988), Johansen & Juselius (1990), MacKinnon (1991) and Reinsel & Ahn (1992).

Application of the likelihood ratio test results in a sequential testing procedure: the test statistic $\mathrm{LR}(r_0, r_1)$ is computed for $r_0 = 0, 1, \ldots, k-1$ until the value of the test statistic exceeds the critical value. Rejection of H_0 then supports a cointegration rank of $r = r_0 + 1$. If the null hypothesis cannot be rejected in the first step ($r_0 = 0$), the test indicates that there is no cointegration and the model should be respecified as a VAR model in first differences. If the null hypothesis is rejected for $r_0 = k - 1$, the test indicates that $y_t \sim I(0)$ and inference should be undertaken in the levels VAR specification.

Least Squares Estimation.

Sims et al. (1990, Theorem 1) have shown that least squares estimation of the levels VAR model yields consistent estimates of the vector autoregressive coefficients. The least squares estimator of $\Pi^{(1)} = [\Pi_1, \ldots, \Pi_p]$ can be computed as in (4.2) and $\hat{\pi} = \mathrm{vec}(\hat{\Pi}^{(1)})$, normalized by $T^{1/2}$, has the same limiting distribution as its maximum likelihood counterpart $\tilde{\pi}$ in (4.41). The covariance matrix $\Sigma_{\hat{\pi}}$ can be consistently estimated as in the stationary, stable case (see Lütkepohl (1991, Proposition 11.3)). Since the least squares estimator does not require information on the number of cointegration relations, estimates of the VAR coefficients can be easily obtained by ordinary least squares if interest does not center on the cointegration relations.

For more details on estimation and inference in cointegrated VAR systems see e.g. Lütkepohl (1991, Chapter 11), Banerjee et al. (1993, Chapter 8), Johansen (1995, Chapter 6) and Hatanaka (1996).

4.2.2 Standard Granger Noncausality

A Wald test for Granger noncausality of y_1 for y_2 in the nonstationary levels VAR model can be based on the same Wald statistic $\mathcal{W}_{\text{GC}}$ as in the stationary VAR model, see (4.8) on page 64. In particular, the Wald statistic can be based upon least squares estimation of the levels VAR model and thus does not need prior information on the order of integration of y_t and the cointegration rank r. However, the limiting distribution of the standard Wald statistic $\mathcal{W}_{\text{GC}}$ may be nonstandard because $\Sigma_{\hat{\pi}}$ is a singular matrix if y_t contains variables which are integrated of order one. The following cases can be distinguished:

If all variables in y_t are stationary, the standard Wald statistic has a limiting $\chi^2(p)$–distribution.

If all variables in y_t are integrated of order one and not cointegrated, the Wald statistic has a nonstandard but nuisance parameter free asymptotic distribution (see Toda & Phillips (1993, Theorem 2)). However, in this case the levels VAR model can be written in first differences without loss of information:

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \ldots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t. \tag{4.42}$$

Since $\Delta y_t \sim I(0)$, estimation and inference in the VAR in first differences can be performed just as in the stationary, stable VAR case. It follows from Proposition 3.4, that y_1 is Granger causal for y_2 if $H_0 : \gamma_{21,1} = \ldots = \gamma_{21,p-1} = 0$ holds. The corresponding Wald statistic has an asymptotic $\chi^2(p-1)$–distribution under this null hypothesis.[6]

If the variables in y_t are cointegrated with cointegration rank $0 < r < k$, the Wald statistic $\mathcal{W}_{\text{GC}}$ has a nonstandard asymptotic distribution with nuisance parameters in general. However, Sims et al. (1990), Toda (1991), Warne (1992a) and Toda & Phillips (1993, Theorem 1; 1994) have shown that $\mathcal{W}_{\text{GC}}$ maintains an asymptotic χ^2–distribution in case of *sufficient cointegration*[7] with respect to variable y_1, i.e. if there is a cointegration relation which includes y_1.

---

[6] Toda & Phillips (1993) find out that the latter test is likely to have higher power in finite samples than a test in a levels VAR model.

[7] If y_1 is a $(k_1 \times 1)$ vector of variables with $k_1 > 1$, then sufficient cointegration holds with respect to y_1 if the k_1 variables in y_1 enter into k_1 different cointegration vectors.

In case of sufficient cointegration with respect to y_1, there exists an alternative representation of the levels VAR model in which the dependent variables are stationary and all the coefficients attached to $y_{1,t-1}, \ldots, y_{1,t-p}$ appear as coefficients of zero–mean stationary regressors. For instance, the error correction model is such an alternative representation: in this model, the estimators of coefficients which belong to the zero–mean stationary regressors converge at a rate of $T^{1/2}$ to a nonsingular normal distribution (see Park & Phillips (1989), Sims et al. (1990) and Dolado & Lütkepohl (1996)). As consequence, the standard Wald statistic $\mathcal{W}_{\text{GC}}$ has a limiting χ^2–distribution under H_0^{GC}.

In a bivariate VAR(p) model with $k_1 = k_2 = 1$ and $k_3 = 0$, a Wald test of Granger noncausality can always be set up in such a way that the Wald statistic has its usual χ^2–distribution if the cointegration rank is known (see Lütkepohl & Reimers (1992a) and Lütkepohl & Reimers (1992b)): if $y_t \sim I(0)$, $\mathcal{W}_{\text{GC}}$ has a limiting $\chi^2(p)$ distribution. The same is true if $y_t \sim I(1)$ and cointegrated, since the condition of sufficient cointegration holds in this case. If however $y_t \sim I(1)$ and not cointegrated, then Granger noncausality should be tested in a VAR($p-1$) model in first differences (see (4.42)). The corresponding Wald statistic is asymptotically $\chi^2(p-1)$–distributed.

To determine the limiting distribution of the Wald statistic $\mathcal{W}_{\text{GC}}$, information on the number of unit roots and the cointegration vectors is needed. In this thesis, it is assumed that $k_1 = k_2 = 1$. Therefore, the condition of sufficient cointegration holds if variable y_1 is cointegrated with variable y_2 and/or the variables in the vector y_3. This can be easily checked with a single equation cointegration test, see e.g. Banerjee et al. (1993, Chapter 7), Hamilton (1994, Chapter 19), Maddala & Kim (1998, Chapter 6).[8] In any case, setting up a Wald test for Granger noncausality in the levels VAR model requires to test for sufficient cointegration in the first step. Since ordinary least squares estimation does not provide information on the cointegration matrix B, a more promising strategy might be to estimate the error correction representation via Johansen's ML procedure, and to test for Granger noncausality within this representation.

[8]Note, that if y_1 is a $(k_1 \times 1)$ vector of variables with $k_1 > 1$, the condition of sufficient cointegration requires a rank test on the cointegration matrix B.

Standard Granger Causality in the Error Correction Representation.

The restrictions for Granger noncausality of y_1 for y_2 in the error correction representation are given in Proposition 3.4 and can be written as linear restrictions on the coefficients in $\psi^\star$, i.e.

$$H_0 : \; R\psi^\star = 0. \tag{4.43}$$

The corresponding Wald statistic

$$\mathcal{W}_{\text{GC}}^{\text{ECM}} \;=\; T(R\tilde{\psi}^\star)' \left(R\tilde{\Sigma}_{\tilde{\psi}^\star} R'\right)^{-1} (R\tilde{\psi}^\star) \tag{4.44}$$

has a limiting $\chi^2(p)$–distribution under the null hypothesis (4.43) if $R\Sigma_{\tilde{\psi}^\star}R'$ is nonsingular and a limiting nonstandard distribution otherwise (see Sims et al. (1990), Toda & Phillips (1993, 1994), Warne (1992a), Warne (1992b) and Warne (1997)). A limiting $\chi^2(p)$–distribution holds in the following two cases:

C1. there is sufficient cointegration with respect to the causal variable y_1, or

C2. the $\Delta y_{2,t}$–equation contains a cointegration vector,

see Toda & Phillips (1993, Theorem 3; 1994, Theorem 2)[9].

Writing the null hypothesis (4.43) as a composite of short–run noncausality (H_0^{SRC}) and long–run noncausality (H_0^{LRC}), i.e.

$$H_0^{\text{LRC}} : \; \Pi(1)_{21} = 0 \qquad \text{and} \qquad H_0^{\text{SRC}} : \; \gamma_{21,1} = \ldots = \gamma_{21,p-1} = 0, \tag{4.45}$$

helps to illustrate conditions C1 and C2:
Under H_0^{SRC}, only coefficients associated with the zero–mean stationary regressors $\Delta y_{t-1}, \ldots, \Delta y_{t-p+1}$ are restricted. It follows that the Wald statistic for a test of the null hypothesis H_0^{SRC} always has a limiting $\chi^2(p-1)$–distribution. The problem of a possibly singular covariance matrix $R\Sigma_{\tilde{\psi}^\star}R'$ is thus caused entirely by the restrictions for long–run noncausality $\Pi(1)_{21} = [AB']_{21} = 0$. Let $A = [a_{ls}]$ and $B = [b_{ls}]$ be $(k \times r)$ matrices with, $l = 1, \ldots, k$, $s = 1, \ldots, r$. The restriction

[9] In general, if y_1 is a $(k_1 \times 1)$ vector of variables with $k_1 > 1$, condition C1 requires that the k_1 variables in y_1 enter into k_1 different cointegration vectors; moreover, if y_2 is a $(k_2 \times 1)$ vector of variables with $k_2 > 1$, condition C2 requires that each of the k_2 $\Delta y_{2,t}$–equations contains a different cointegration vector.

$\Pi(1)_{21} = [AB']_{21} = a_{21}b_{11} + \ldots + a_{2r}b_{r1} = 0$ is a nonlinear function of the coefficients in A and B. Assume the most simple case of just one cointegrating vector ($r = 1$), then the restriction $\Pi(1)_{21} = 0$ is fulfilled if either

(i) $a_{21} = 0$ or
(ii) $b_{11} = 0$ or
(iii) $a_{21} = b_{11} = 0$.

However, case (iii) contains a redundant restriction[10] such that the Jacobian matrix of the nonlinear restriction will not have full rank under H_0^{LRC}. As consequence, the Wald statistic $\mathcal{W}_{\text{GC}}^{\text{ECM}}$ will have a nonstandard limiting distribution in case (iii). However, conditions C1 and C2 of Toda & Phillips simply rule out the parameter values in (iii): Under condition C1, at least one element $b_{1s} \neq 0$ for $s \in \{1, \ldots, r\}$; under condition C2, at least one element $a_{2s} \neq 0$ for $s \in \{1, \ldots, r\}$. In both cases, the Jacobian matrix will have full row rank under H_0^{LRC}. The limiting χ^2–distribution of $\mathcal{W}_{\text{GC}}^{\text{ECM}}$ then follows from the limiting normal distributions of $\sqrt{T}\text{vec}(\tilde{A} - A)$, the limiting mixed Gaussian distribution of $\sqrt{T}\text{vec}(\tilde{B} - B)$ and the limiting normal distribution of $\sqrt{T}\text{vec}(\tilde{\Psi} - \Psi)$ (see Johansen (1995, Chapter 7)).

Setting up a standard Wald test of Granger noncausality in the error correction representation requires to test conditions C1 and C2 in the first step. This leads to a sequential testing procedure:
For instance, a Wald test of the null hypothesis $H_0^A : a_{21} = \ldots = a_{2r} = 0$ may be undertaken in the first step. The corresponding Wald statistic has a limiting $\chi^2(r)$–distribution. If H_0^A is rejected, the Wald statistic $\mathcal{W}_{\text{GC}}^{\text{ECM}}$ can be computed in the second step and compared to an $\alpha 100\%$ critical value of a $\chi^2(p)$–distribution. If H_0^A cannot be rejected, y_1 is long–run noncausal for y_2 and it suffices to test for short–run noncausality in the second step. The corresponding Wald statistic has a limiting $\chi^2(p-1)$–distribution under H_0^{SRC}.
Alternatively, one may start with a Wald test of the null hypothesis $H_0^B : b_{11} = \ldots = b_{1r} = 0$. If H_0^B is rejected, the composite null hypothesis (4.43) has to be tested in the

[10] See Example 4.1 for a restriction with similar structure.

second step. If H_0^B cannot be rejected, H_0^{SRC} is tested in the second step.[11]
A drawback of both sequential testing procedures is that the overall significance level cannot be controlled as for instance in case of induced tests.[12] Simulation results on the actual size and the local power properties of both sequential testing procedures are given in Toda & Phillips (1994) for a trivariate VAR(1) model.

Like Wald tests for Granger noncausality in the levels VAR model, Wald tests for Granger noncausality in the error correction representation require information on the true cointegration rank. In practice, this information is unknown in general so that causality tests have to be based on pretests for unit roots, cointegration and the nature of cointegration. Since the latter tests may have low power in small samples, causality tests may suffer from large pretest biases and low power, too (see Sims et al. (1990)). Moreover, the simulation results of Toda & Phillips (1994) show that size distortions may arise if the analysis is based on an incorrect cointegration rank. It would be therefore favorable if such pretests could be avoided altogether. Some work in this direction has been undertaken by Phillips (1995), Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996):

Alternative Testing Strategies

Based on earlier work of Phillips & Hansen (1990), Phillips (1995) proposes a fully modified (FM–VAR) estimator to obtain estimates of the levels VAR coefficient matrices $\Pi = [\Pi_1, \ldots, \Pi_p]$ without any information on the number of unit roots and the cointegration rank. Without going into too much detail, the idea of the FM–VAR estimator shall be briefly outlined: the FM–VAR estimator

$$\hat{\Pi}_{\text{FM}} \quad = \quad (y^+ Z' - T\hat{\Delta}_{0Z}^+)(ZZ')^{-1}. \tag{4.46}$$

[11] Note, that this procedure holds only for $k_1 = k_2 = 1$. If causality between vectors of variables is under study $(k_1, k_2 > 1)$, condition 1 requires to test the row rank of the k_1 rows of the cointegration matrix B which belong to $y_{1,t-1}$ while condition 2 requires to test the row rank of the k_2 rows of the loading matrix A which appear in the equations of $\Delta y_{2,t}$. See Mosconi & Giannini (1992) for suggestions on a testing procedure.

[12] A sequential test is called an induced test if rejection of a single null hypothesis H_0^i for some $i = 1, \ldots, n$ implies rejection of the overall null hypothesis $H_0 = H_0^1 \cap H_0^2 \cap \ldots \cap H_0^n$. In the present case, these conditions are not met because rejection of H_0^A or H_0^B does not imply rejection of the composite null hypothesis H_0^{ECM}. See Savin (1984), Krämer & Sonnberger (1986).

can be interpreted as a least squares estimator which corrects the term yZ' in (4.2), p. 63, for endogeneities between y and Z due to cointegration and for serial correlation between the nonstationary part of the regressors in Z and the disturbance vector u: In (4.46), y^+ denotes y corrected for endogeneity and $\hat{\Delta}^+_{0Z}$ is the serial correlation correction term (see Phillips (1995, formula (7))). If the Wald statistic $\mathcal{W}_{\text{GC}}$ is based on the FM–VAR estimator, it follows from Theorem 4.5 and the rank condition (RK) of Phillips (1995), that this Wald statistic has a limiting $\chi^2(p)$–distribution if either all variables in y_t are stationary of if there is sufficient cointegration with respect to y_1. If these conditions do not hold, the Wald statistic has a limiting distribution which is degenerate for a lag order $p = 1$ and which is a mixture of χ^2–variates for lag orders $p > 1$. In the latter case, the $\alpha 100\%$–critical value of the true limiting distribution is bounded between the $\alpha 100\%$–critical values of a $\chi^2(p-1)$– and a $\chi^2(p)$–distribution. Hence, the FM–VAR based Wald test is a conservative test which has asymptotic size less than or equal to $\alpha 100\%$ if used with an $\alpha 100\%$–critical value of a $\chi^2(p)$–distribution. Note, that the FM–VAR based Wald test works only for lag orders $p > 1$. For simulation results on the empirical size and local power properties, see Yamada & Toda (1997, 1998).

Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) propose another testing strategy which also does not require information on the number of unit roots and the cointegration rank r: given that the vector of variables y_t is at most integrated of order d, that p is the true lag order of the VAR model and that $d < p$ holds, the authors propose to fit a VAR$(p+d)$ model to the true VAR(p) model (see also Choi (1993)). Least squares estimation yields estimates $\hat{\Pi}_i$ of the coefficient matrices $\Pi_1, \ldots, \Pi_p, \ldots, \Pi_{p+d}$. The authors show that if restrictions are tested on only the first p coefficient matrices $(\Pi_1, \ldots, \Pi_p)$, the VAR$(p+d)$ model can be transformed in such a way that only coefficients of stationary regressors are restricted in the transformed model (see Sims et al. (1990), Toda & Yamamoto (1995, Theorem 1), Dolado & Lütkepohl (1996, Theorem 1)). A little example may help illustrate the idea:

Example 4.2: VAR(2) Model.

Assume that the variables in y_t are at most $I(d)$ with $d = 1$ and are generated by the

following VAR(2) model:

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + u_t.$$

Of course, the VAR(3) model

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \Pi_3 y_{t-3} + u_t \tag{4.47}$$

with $\Pi_3 = 0$ describes the data generating process equally well. The latter model can be written as

$$\begin{aligned} \Delta y_t &= (\Pi_1 - I_k)\Delta y_{t-1} + (\Pi_1 + \Pi_2 - I_k)\Delta y_{t-2} \qquad (4.48) \\ &- (\Pi_1 + \Pi_2 + \Pi_3 - I_k) y_{t-3} + u_t, \\ &= \Gamma_1^\star \Delta y_{t-1} + \Gamma_2^\star \Delta y_{t-2} - \Pi(1) y_{t-3} + u_t. \qquad (4.49) \end{aligned}$$

In the levels VAR model (4.47), Granger noncausality of y_1 for y_2 holds if $\pi_{21,1} = \pi_{21,2} = 0$. These restrictions translate into equivalent restrictions $\gamma^\star_{21,1} = \gamma^\star_{21,2} = 0$ in the transformed model (4.49). The latter restrictions are now associated with the zero–mean stationary regressors $\Delta y_{t-1}, \Delta y_{t-2}$. As consequence, a Wald test of the latter restrictions follows asymptotically a χ^2–distribution.

To set up a Wald test, the levels VAR model (4.47) is estimated by OLS. This yields the least squares estimator $\hat{\Pi} = [\hat{\Pi}_1,\ \hat{\Pi}_2,\ \hat{\Pi}_3]$ and $\hat{\pi} = \text{vec}(\hat{\Pi})$, with covariance matrix $\Sigma_{\hat{\pi}}$. Let $\pi^\star = \text{vec}\,[\Pi_1,\ \Pi_2]$ and $\hat{\pi}^\star = \text{vec}\left[\hat{\Pi}_1,\ \hat{\Pi}_2\right]$, then

$$\sqrt{T}(\hat{\pi}^\star - \pi^\star) \xrightarrow{d} N(0, \Sigma_{\hat{\pi}^\star}), \tag{4.50}$$

with $\Sigma_{\hat{\pi}^\star}$ a nonsingular covariance matrix which is the $(k^2p \times k^2p)$ upper left–hand submatrix of $\Sigma_{\hat{\pi}}$.

Furthermore, let $\hat{\Sigma}_{\hat{\pi}^\star}$ denote a consistent estimator of $\Sigma_{\hat{\pi}^\star}$ and let R be the $(p \times k^2p)$ matrix defined in (4.5) which picks out of $\hat{\pi}^\star$ the coefficients $\pi_{21,1}$ and $\pi_{21,2}$. Then under the null hypothesis

$$H_0 : R\pi^\star = 0,$$

the Wald statistic

$$\mathcal{W}^\star_{\text{GC}} = T\,(R\hat{\pi}^\star)' \left[R\hat{\Sigma}_{\hat{\pi}^\star} R'\right]^{-1} (R\hat{\pi}^\star) \xrightarrow{d} \chi^2(p) \tag{4.51}$$

(see Dolado & Lütkepohl (1996, Theorem 1).

The Wald statistic $\mathcal{W}^{\star}_{\mathrm{GC}}$ is consistent, has correct asymptotic size and is easy to set up. However, the inefficiency introduced by overfitting the true VAR order may result in a loss in power. In practice, $d = 1$ should yield a good description of many time series. Adding an extra lag to the true lag order is likely to reduce the power the higher the dimension of the VAR model and the smaller the true lag order. On the other hand, the loss should not be substantial for low–dimensional VAR models, in particular if the true lag order is high. The simulation results of Dolado & Lütkepohl (1996), Zapata & Rambaldi (1997) and Yamada & Toda (1998) support this reasoning.

4.2.3 Extended Granger Noncausality

A test of the null hypothesis H_0^{EGC} can be based on the same Wald statistic $\mathcal{W}_{\mathrm{EGC}}(\tilde{h})$ as in the stationary VAR(p) model (see page 66). However, section 4.1.3 has illustrated that the Wald statistic $\mathcal{W}_{\mathrm{EGC}}(\tilde{h})$ may not have a limiting $\chi^2(\tilde{h}p)$–distribution for some parameter values under H_0^{EGC} if these parameter values imply a reduced row rank of the Jacobian matrix $\partial g(\pi)/\partial \pi'$ and hence a singular covariance matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi')\Sigma_{\hat{\pi}}(\partial g(\pi)'/\partial \pi)$.

In the nonstationary VAR model, $\Sigma_{\hat{\pi}}$ is singular. Therefore, $\Sigma_{g(\hat{\pi})}$ may be singular even if the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has full rank under H_0^{EGC}. Section 4.2.2 has shown that the Wald statistic $\mathcal{W}_{\mathrm{GC}}$ maintains a limiting $\chi^2(p)$–distribution if there is sufficient cointegration with respect to y_1. However, the condition of sufficient cointegration cannot guarantee that $\mathcal{W}_{\mathrm{EGC}}(\tilde{h})$ has a limiting $\chi^2(\tilde{h}p)$–distribution under H_0^{EGC}:

It has been shown in Section 4.2.2, that a Wald test of standard Granger causality maintains its limiting $\chi^2(p)$–distribution if there is sufficient cointegration with respect to y_1. But the restrictions in Example 4.1 show that Granger noncausality of y_1 for y_2 at higher forecast horizons restricts coefficients associated with $y_{1,t-i}$ as well as $y_{3,t-i}$ (see page 67). The condition of sufficient cointegration with respect to y_1 ensures only that the restrictions on coefficients associated with $y_{1,t-i}$ in the $y_{2,t}$–equation can be reformulated as restrictions on coefficients of zero–mean stationary regressors. Hence, even if sufficient cointegration with respect to y_1 holds, the restrictions for Granger

noncausality at higher forecast horizons may still restrict coefficients attached to nonstationary regressors, leading to a nonstandard asymptotic distribution of $\mathcal{W}_{\text{EGC}}(\tilde{h})$.

If $y_t \sim I(1)$ and not cointegrated, the nonstationary levels VAR(p) model can be written as a stationary VAR($p-1$) model in first differences (see (4.42)) and estimated by ordinary least squares. A Wald test that y_1 is never causal for y_2 can then be set up just as described in Section 4.1.3. However, this procedure requires a pretest on the number of unit roots and the cointegration rank and applies only in the special case where a cointegration rank $r = 0$.

In contrast, the problem of a singular covariance matrix $\Sigma_{\hat{\pi}}$ can be solved without any information on the number of unit roots and the nature of cointegration if a Wald test that y_1 is never Granger causal for y_2 is set up along the lines of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996). The Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ can be computed just as in the stationary, stable VAR case if a VAR($p+d$) model is estimated, but if only the coefficients of the first p estimated matrices $\hat{\Pi}_1, \ldots, \hat{\Pi}_p$ are used in the computations. However, the Wald statistic may continue to have a nonstandard asymptotic distribution if the Jacobian matrix $\partial g(\pi)/\partial \pi'$ in (4.11), page 66, has reduced row rank under the null hypothesis H_0^{EGC}. Solutions for this problem are discussed in the next chapter.

The FM–VAR estimator cannot be regarded as an alternative to the strategy of overfitting the true lag length because the covariance matrix of the FM–VAR estimator is singular. While Phillips shows for the case of linear restrictions that the FM–VAR based standard Wald test is conservative when used with an $\alpha 100\%$ critical value of a χ^2–distribution, this may not hold for nonlinear restrictions (see Phillips (1995), Yamada & Toda (1998)).

4.2.4 Impulse Response Analysis

Impulse responses can be computed just as in the stationary levels VAR model. However, restricting impulse response coefficient to zero imposes nonlinear restrictions on

the VAR coefficients (see Section 4.1.4). These restrictions are similar to the nonlinear restrictions which characterize Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$. Therefore, the problems outlined in Section 4.2.3 carry over to impulse response analysis: the Wald statistic $\mathcal{W}_{\text{IR}}(\bar{h})$ will have a nonstandard limiting distribution in general if unit roots and cointegration are present.

A Wald test that y_1 is never causal for y_2 in terms of impulse response analysis can also be set up along the lines of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) as described above. This solves the problem of a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ due to nonstationary regressors but cannot solve the problems outlined in Section 4.1.4.

4.3 Summary

Standard Wald tests of causality may be fraught with the possibility of nonstandard asymptotic distributions if restrictions are nonlinear or if variables are nonstationary: In stationary VAR models, Wald tests of Granger noncausality at forecast horizon $h = 1$ do not suffer from any problems while Wald tests of Granger noncausality at all forecast horizons $h \geq 1$ and Wald tests of significance of all impulse responses do because of the nonlinearity of the restrictions under the respective null hypotheses.
If the k–dimensional vector of variables y_t is integrated of order one and cointegrated, the singularity of $\Sigma_{\hat{\pi}}$, the covariance matrix of the levels VAR coefficients, creates another source which may lead to a nonstandard asymptotic distribution of the Wald statistics $\mathcal{W}_{\text{EGC}}$ and $\mathcal{W}_{\text{IR}}$. In the latter case, even the Wald statistic ($\mathcal{W}_{\text{GC}}$) for a test of Granger noncausality at forecast horizon $h = 1$ may no longer be asymptotically $\chi^2(p)$–distributed under H_0.
Different solutions to the latter problem have been presented e.g. by Sims et al. (1990), Toda (1991), Mosconi & Giannini (1992), Toda & Phillips (1993, 1994), Phillips (1995), Warne (1992a, b; 1997), Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996). In particular, Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) have presented a solution which guarantees a nonsingular covariance matrix $\Sigma_{\hat{\pi}}$, independent of the nature of the restrictions. However, even if $\Sigma_{\hat{\pi}}$ is nonsingular, the asymptotic distribu-

tion of the Wald statistic may be nonstandard if restrictions are nonlinear under the null hypothesis.

4.4 Concluding Remarks

The analysis in this chapter has been based on a known lag order p. In general, p is unknown but can be determined consistently by the Hannan–Quinn (HQ) or the Schwarz (SC) criterion. Consistency of these order selection criteria holds also for the nonstationary VAR models discussed here (Paulsen (1984)). A description of these criteria can be found in Lütkepohl (1991, Chapters 4 and 11). The properties of the Wald tests discussed in this chapter hold under the assumption that the lag order p is either known or estimated correctly. Misspecification of the lag order may result in size distortions and loss in power.[13] For instance, overfitting the true VAR order may lead to a reduced row rank of the matrix of first order partial derivatives (see Warne (1992a, 1997)) such that a Wald test for Granger noncausality may no longer have a limiting χ^2–distribution. Note, however, that problems due to under– or overestimation of the true lag order p have not been considered here.

Throughout this thesis, only VAR models of finite lag order p are considered. In practice, the true lag order may be infinite. In this case, a VAR model of finite lag order p may still be fitted to the data under suitable assumptions about the rate at which the lag order p converges to infinity with increasing sample size. Lütkepohl (1996b), Lütkepohl & Poskitt (1996) and Lütkepohl & Saikkonen (1997) show how the standard Wald statistic should be modified under this assumption to test for zero impulse responses. Since the restrictions for zero impulse responses are similar to those for Granger noncausality at all forecast horizons, results may carry over to extended Granger causality tests under suitable modifications.

Third, deterministic terms have been ignored in the regression equations. While deterministic terms do not play a role in the definition of Granger causality, extended Granger causality or impulse response analysis, they do play an important role in the estimation of nonstationary VAR models. Moreover, unit root and cointegration tests

[13] Braun & Mittnik (1993) study the consequences of a misspecified lag order for impulse response analysis.

and the corresponding critical values depend on the deterministic terms. However, inclusion of deterministic terms does not change the results presented in this chapter.

Chapter 5

Causality in VAR Models: Alternative Testing Strategies.

This chapter concentrates on extended Granger causality and discusses solutions to overcome the problem of a possibly nonstandard asymptotic distribution of the Wald statistic. In particular, two modified Wald statistics, namely a randomized Wald statistic and a Wald statistic with generalized inverse, are presented which always have a known asymptotic distribution under the null hypothesis H_0^{EGC} that y_1 is never Granger causal for y_2. These modified Wald statistics can also be used for a test that all responses of y_2 to a one–time, one–unit shock in y_1 are zero. Moreover, suggestions on alternative testing strategies are given which are not in the spirit of Wald tests.

5.1 Stationary, Stable VAR Models

The restrictions for Granger noncausality of y_1 for y_2 at all forecast horizons are given in (4.9), see page 65. Let $g(\hat{\pi})$ denote the least squares estimator of the vector of restrictions $g(\pi)$, then under the assumptions of Section 4.1.3,

$$\sqrt{T}(g(\hat{\pi}) - g(\pi)) \xrightarrow{d} N(0, \Sigma_{g(\hat{\pi})}), \tag{5.1}$$

if $\partial g(\pi)/\partial \pi'$ exists and is nonzero at π (see Serfling (1980, p. 122-124)). However, $\Sigma_{g(\hat{\pi})}$ may be singular under the null hypothesis $H_0 : g(\pi) = 0$ if some elements of $g(\hat{\pi})$ converge towards their theoretical value of zero at a rate faster than $T^{-1/2}$. In this

case, the standard Wald statistic

$$\mathcal{W} = T g(\hat{\pi})' \hat{\Sigma}^{-1}_{g(\hat{\pi})} g(\hat{\pi}) \tag{5.2}$$

may no longer have an asymptotic χ^2–distribution under the null hypothesis.

In what follows, modifications of the standard Wald tests will be presented: for instance the randomized Wald test of Lütkepohl & Burda (1997, Proposition 1) works via adding some random noise to the estimator $g(\hat{\pi})$, thus slowing down the convergence rate of $g(\hat{\pi})$.[1]

A different starting point is to find a reduced rank estimator of $\Sigma_{g(\hat{\pi})}$: for instance, the Wald test with generalized inverse (Lütkepohl & Burda (1997, Proposition 2)) is based on estimating the rank of $\Sigma_{g(\hat{\pi})}$ and then replacing the inverse $\hat{\Sigma}^{-1}_{g(\hat{\pi})}$ in (5.2) by a generalized inverse $\hat{\Sigma}^{+}_{g(\hat{\pi})}$.

5.1.1 A Randomized Wald Test

To illustrate the general idea of the randomized Wald test first, consider the null hypothesis

$$H_0 : \; g(\pi) = 0, \tag{5.3}$$

with $g(\pi) = [g_1(\pi)', g_2(\pi)']$ an $(r_g \times 1)$ vector consisting of an $(r_1 \times 1)$ subvector of linear restrictions $g_1(\pi)$ and an $(r_2 \times 1)$ subvector of nonlinear restrictions $g_2(\pi)$ with $r_1, r_2 > 0$ and $r_g = r_1 + r_2$. Let $g(\hat{\pi})$ denote a consistent estimator of $g(\pi)$ with limiting normal distribution in (5.1) and $\Sigma_{g(\hat{\pi})}$ possibly singular under H_0. The idea of the randomized Wald test is to render the estimation of $g(\hat{\pi})$ less efficient by adding a zero–mean random vector w_ξ to the set of r_g restrictions under H_0:

Let w_ξ be an $(r_g \times 1)$ vector independent of $\hat{\pi}$ and normally distributed with mean zero and covariance matrix $\xi \Sigma_w$ for some small $\xi > 0$. Then

$$\sqrt{T}\left(g(\hat{\pi}) + \frac{w_\xi}{\sqrt{T}}\right) \xrightarrow{d} N(0, (\Sigma_{g(\hat{\pi})} + \xi \Sigma_w))$$

under H_0. If Σ_w is nonsingular, $\Sigma_{g(\hat{\pi})} + \xi \Sigma_w$ will be a nonsingular matrix, independent of the rank of $\Sigma_{g(\hat{\pi})}$.

---

[1] An extension of this idea to distance tests is studied in Bolfarine et al. (2001).

To keep the loss in efficiency small, no superfluous random noise should be added to $g(\pi)$. In particular, since the first r_1 restrictions in $g(\pi)$ are linear, the upper left–hand $(r_1 \times r_1)$ submatrix of $\Sigma_{g(\hat{\pi})}$ is nonsingular and $\mathrm{rk}(\Sigma_{g(\hat{\pi})}) \geq r_1$. Therefore, the first r_1 elements of w_ξ should be set equal to zero with zero variance, such that

$$\Sigma_w = \begin{bmatrix} 0 & 0 \\ 0 & \Sigma_w(r_2) \end{bmatrix},$$

with $\Sigma_w(r_2)$ a $(r_2 \times r_2)$ nonsingular covariance matrix. Let

$$\Sigma_{g(\hat{\pi})} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

with Σ_{11} a $(r_1 \times r_1)$ nonsingular submatrix belonging to the first r_1 restrictions in $g_1(\hat{\pi})$. Then it follows that $(\Sigma_{g(\hat{\pi})} + \xi\Sigma_w)$ is nonsingular, and the modified Wald statistic

$$\begin{aligned} \mathcal{W}^{(\mathrm{ran})} &= T\left(g(\hat{\pi}) + \frac{w_\xi}{\sqrt{T}}\right)' \left(\hat{\Sigma}_{g(\hat{\pi})} + \xi\Sigma_w\right)^{-1} \left(g(\hat{\pi}) + \frac{w_\xi}{\sqrt{T}}\right) \\ &\xrightarrow{d} \chi^2(r_g) \end{aligned}$$

under H_0 also in those cases where $r_1 \leq \mathrm{rk}(\Sigma_{g(\hat{\pi})}) < r_g$.

Suppose that $\hat{w}_\xi \sim N(0,\ \xi\hat{\Sigma}_w)$ is generated just as w_ξ but with covariance matrix $\xi\hat{\Sigma}_w$. Then if $\hat{\Sigma}_w$ is a consistent estimator of Σ_w such that $\hat{w}_\xi \xrightarrow{d} w_\xi \sim N(0, \xi\Sigma_w)$ which is independent of $\hat{\pi}$ by assumption, it follows from basic asymptotic results that also the Wald statistic

$$\begin{aligned} \mathcal{W}^{(\xi)} &= T\left(g(\hat{\pi}) + \frac{\hat{w}_\xi}{\sqrt{T}}\right)' \left(\hat{\Sigma}_{g(\hat{\pi})} + \xi\hat{\Sigma}_w\right)^{-1} \left(g(\hat{\pi}) + \frac{\hat{w}_\xi}{\sqrt{T}}\right) \\ &\xrightarrow{d} \chi^2(r_g) \end{aligned} \tag{5.4}$$

(see Lütkepohl & Burda (1997, Proposition 1)).

For a value $\xi = 0$, the randomized Wald statistic $\mathcal{W}^{(\xi)}$ reduces to the standard Wald statistic. For $\xi > 0$, adding $\hat{w}_\xi$ to the estimator $g(\hat{\pi})$ will reduce the power of the Wald test. The loss in power hinges on the size of the covariance matrix $\hat{\Sigma}_w$ relative to the size of $\hat{\Sigma}_{g(\hat{\pi})}$. To minimize the loss, the variability of the noise should be related to the variability of the estimator: by choosing a small ξ value, the noise can be made arbitrarily small. Note, however, that the modified Wald test has correct asymptotic size for any $\xi \geq 0$.

Extended Granger Causality

The principle of the randomized Wald test will be illustrated for the null hypothesis that y_1 is never Granger causal for y_2 (see (4.9) on page 65). Under H_0^{EGC}, $\tilde{h}p$ restrictions are imposed on the pk^2 VAR coefficients. Thereby, the first p restrictions are linear and the remaining $(\tilde{h}-1)p$ restrictions are nonlinear. Hence, $r_g = \tilde{h}p = r_1 + r_2$ with $r_1 = p$ and $r_2 = (\tilde{h}-1)p$.

The covariance matrix

$$\begin{aligned}\Sigma_{g(\hat{\pi})} &= \frac{\partial g(\pi)}{\partial \pi'}\Sigma_{\hat{\pi}}\frac{\partial g(\pi)'}{\partial \pi} \\ &= (I_{\tilde{h}} \otimes R)\left(\frac{\partial \boldsymbol{\pi}^{(\tilde{h})}}{\pi'}\Sigma_{\hat{\pi}}\frac{\partial \boldsymbol{\pi}^{(\tilde{h})'}}{\pi}\right)(I_{\tilde{h}} \otimes R') \\ &= (I_{\tilde{h}} \otimes R)\Sigma_{\hat{\pi}}(\tilde{h})(I_{\tilde{h}} \otimes R'),\end{aligned}$$

with

$$\Sigma_{\hat{\pi}}(\tilde{h}) = \frac{\partial \boldsymbol{\pi}^{(\tilde{h})}}{\pi'}\Sigma_{\hat{\pi}}\frac{\partial \boldsymbol{\pi}^{(\tilde{h})'}}{\pi}.$$

Let $\hat{\Sigma}_{\hat{\pi}}(\tilde{h})$ denote a consistent estimator of $\Sigma_{\hat{\pi}}(\tilde{h})$ with

$$\hat{\Sigma}_{\hat{\pi}}(\tilde{h}) = \frac{\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})}}{\partial \pi'}\hat{\Sigma}_{\hat{\pi}}\frac{\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})'}}{\partial \pi}$$

and $\partial \hat{\boldsymbol{\pi}}^{(\tilde{h})'}/\partial\pi$ computed just like

$$\frac{\partial \boldsymbol{\pi}^{(\tilde{h})'}}{\partial \pi} = \left[\begin{array}{cccc} I_{k^2p} & \frac{\partial \pi^{(2)'}}{\partial \pi} & \cdots & \frac{\partial \pi^{(\tilde{h})'}}{\partial \pi} \end{array}\right],$$

but with the true VAR parameters replaced by their least squares estimates. Then the following proposition can be given:

Proposition 5.1: (Lütkepohl & Burda (1997, Proposition 1 and Corollary 1)).
Let

$$\Sigma_w = \left[\begin{array}{cc} 0 & 0 \\ 0 & I_{\tilde{h}-1} \otimes \operatorname{diag}(R\Sigma_{\hat{\pi}}R') \end{array}\right] \quad (\tilde{h}p \times \tilde{h}p), \tag{5.5}$$

and

$$\hat{\Sigma}_w = \left[\begin{array}{cc} 0 & 0 \\ 0 & I_{\tilde{h}-1} \otimes \operatorname{diag}(R\hat{\Sigma}_{\hat{\pi}}R') \end{array}\right]$$

be a consistent estimator of Σ_w. Then it follows from (5.4) that the Wald statistic

$$\begin{aligned}\mathcal{W}_{\text{EGC}}^{(\xi)}(\tilde{h}) &= T\left((I_{\tilde{h}}\otimes R)\,\hat{\boldsymbol{\pi}}^{(\tilde{h})}+\frac{\hat{w}_\xi}{\sqrt{T}}\right)' \\ &\quad\times\left[(I_{\tilde{h}}\otimes R)\,\hat{\Sigma}_{\hat{\pi}}(\tilde{h})\,(I_{\tilde{h}}\otimes R')+\xi\hat{\Sigma}_w\right]^{-1} \\ &\quad\times\left((I_{\tilde{h}}\otimes R)\,\hat{\boldsymbol{\pi}}^{(\tilde{h})}+\frac{\hat{w}_\xi}{\sqrt{T}}\right) \\ &\xrightarrow{d}\ \chi^2\left(\tilde{h}p\right) \qquad (5.6)\end{aligned}$$

under H_0^{EGC}.

Proof: Since plim $\hat{\Sigma}_w=\Sigma_w$, we have $\hat{w}_\xi \xrightarrow{d} w_\xi \sim N(0,\xi\Sigma_w)$ which is independent of $\hat{\pi}$ by assumption. Moreover, regularity of $(I_{\tilde{h}}\otimes R)\,\Sigma_{\hat{\pi}}(\tilde{h})\,(I_{\tilde{h}}\otimes R')+\xi\Sigma_w$ follows because $\Sigma_{\hat{\pi}}$ and hence $R\Sigma_{\hat{\pi}}R'$, the upper left–hand $(p\times p)$ submatrix of $(I_{\tilde{h}}\otimes R)\,\Sigma_{\hat{\pi}}(\tilde{h})\,(I_{\tilde{h}}\otimes R')$, is nonsingular. Furthermore, the lower right–hand $((\tilde{h}-1)p\times(\tilde{h}-1)p)$ submatrix of $\xi\Sigma_w$ is a diagonal matrix with positive diagonal elements which is also nonsingular. Thus Proposition 5.1 follows from (5.4) and basic asymptotic results. □

Note, that $R\Sigma_{\hat{\pi}}R'$ is the upper left–hand $(p\times p)$ submatrix of $\Sigma_{g(\hat{\pi})}$, which is the nonsingular covariance matrix of the first p linear restrictions in $g(\pi)$. The dependence of Σ_w on $R\Sigma_{\hat{\pi}}R'$ in (5.5) serves two purposes: first, by making the variability of the noise dependent on the variability of the estimator, it shall be avoided that "too much noise is added in relation to the estimated variance" (see Lütkepohl & Burda (1997)). Second, this choice guarantees that the lower $(r_2\times r_2)$ block of $\xi\Sigma_w$ is nonsingular. The following examples shall help to illustrate Proposition 5.1:

Example 5.1: Trivariate VAR(1) Model.

The null hypothesis that y_1 is never causal for y_2 imposes $\tilde{h}p=2$ restrictions. These have been already explored in Examples 3.4 and 4.1:

$$H_0^{\text{EGC}}:\ g(\pi) = \begin{bmatrix}\pi_{21,1}\\ \pi_{21,1}\pi_{11,1}+\pi_{22,1}\pi_{21,1}+\pi_{23,1}\pi_{31,1}\end{bmatrix}=0.$$

Since the first restriction is linear, noise will only be added to the second restriction. A value $\hat{w}_\xi$ is found by drawing randomly from a normal distribution with mean zero and covariance matrix $\xi R\hat{\Sigma}_{\hat{\pi}}R'$. For the present case, $R\hat{\Sigma}_{\hat{\pi}}R'=\hat{\sigma}^2_{\hat{\pi}_{21}}$ is just the variance

of $\hat{\pi}_{21}$, and $\hat{w}_\xi$ can be generated by drawing from a standard normal distribution and multiplying the value with $\sqrt{\xi\hat{\sigma}^2_{\hat{\pi}_{21}}}$. Thereby, a value $\xi > 0$ can be chosen arbitrarily.

Example 5.2: Trivariate VAR(2) Model.

The null hypothesis that y_1 is never causal for y_2 imposes $\tilde{h}p = 6$ restrictions:

$$H_0^{\text{EGC}}: \; g(\pi) \;=\; (I_3 \otimes R)\boldsymbol{\pi}^{(3)} = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,2} \\ \pi^{(2)}_{21,1} \\ \pi^{(2)}_{21,2} \\ \pi^{(3)}_{21,1} \\ \pi^{(3)}_{21,2} \end{bmatrix} = 0,$$

with $R = I_2 \otimes [0,\, 1,\, 0,\, \ldots,\, 0]$ a (2×18) matrix. For the present example, $\hat{\Sigma}_w$ can be partitioned as

$$\hat{\Sigma}_w \;=\; \begin{bmatrix} 0_{2\times 2} & 0_{2\times 4} \\ 0_{4\times 2} & \left(I_2 \otimes \operatorname{diag}(R\hat{\Sigma}_{\hat{\pi}}R')\right) \end{bmatrix}$$

with

$$\operatorname{diag}(R\hat{\Sigma}_{\hat{\pi}}R') \;=\; \begin{bmatrix} \hat{\sigma}^2_{\hat{\pi}_{21,1}} & 0 \\ 0 & \hat{\sigma}^2_{\hat{\pi}_{21,2}} \end{bmatrix},$$

and $\hat{\sigma}^2_{\hat{\pi}_{21,i}}$ the variance of the estimator $\hat{\pi}_{21,i}$ for $i = 1, 2$.

Let

$$P \;=\; \begin{bmatrix} 0_{2\times 2} & 0_{2\times 4} \\ 0_{4\times 2} & P_{22} \end{bmatrix}$$

with

$$P_{22} \;=\; I_2 \otimes \begin{bmatrix} \sqrt{\hat{\sigma}^2_{\hat{\pi}_{21,1}}} & 0 \\ 0 & \sqrt{\hat{\sigma}^2_{\hat{\pi}_{21,2}}} \end{bmatrix},$$

so that $\hat{\Sigma}_w = PP'$. The vector $\hat{w}_\xi$ is found by right–multiplying a (6×1) vector drawn from a standard normal distribution with $\sqrt{\xi}P$.

Impulse Response Analysis

The randomized Wald test can also be used to test for noncausality of y_1 for y_2 in terms of impulse response analysis under slight modifications. These are given in the following corollary:

Corollary 5.1:
A test that all responses of y_2 to a one–time, one–unit shock in y_1 are zero can be performed as in (5.6), but with the selector matrix R replaced by $\bar{R}$ (see page 75) and with the upper bound $\tilde{h} = pk_3 + 1$ replaced by $\bar{h} = p(k-1)$. The modified Wald statistic then has a limiting $\chi^2(\bar{h})$–distribution under H_0^{IR}.

5.1.2 A Wald Test with Generalized Inverse

Problems with the standard Wald test arise if $\Sigma_{g(\hat{\pi})}$ is singular so that the inverse does not exist. Following Andrews (1987) and Vuong (1987), one solution is to derive a generalized inverse $\Sigma^+_{g(\hat{\pi})}$ of $\Sigma_{g(\hat{\pi})}$ such that $\text{rk}(\Sigma^+_{g(\hat{\pi})}) = \text{rk}(\Sigma_{g(\hat{\pi})}) = r_c$, and then use a consistent estimator of $\Sigma^+_{g(\hat{\pi})}$ to set up the Wald statistic. It then follows from Vuong (1987, Theorem 1(ii)), that the modified Wald statistic

$$\mathcal{W}^+ = Tg(\hat{\pi})'\hat{\Sigma}^+_{g(\hat{\pi})}g(\hat{\pi}) \xrightarrow{d} \chi^2(r_c) \tag{5.7}$$

under the null hypothesis. Since the true rank of $\Sigma_{g(\hat{\pi})}$ is unknown in practice, this solution still leaves open the problem of finding a suitable reduced rank estimator in the first step.

Lütkepohl & Burda (1997, Proposition 2) propose to decompose the $(r_g \times r_g)$ covariance matrix of the estimator $g(\hat{\pi})$ such that $\Sigma_{g(\hat{\pi})} = V\Lambda V'$. Thereby, Λ denotes a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{r_g})$ with the eigenvalues of $\Sigma_{g(\hat{\pi})}$ on its main diagonal, and V is the matrix of corresponding eigenvectors. Assume that the eigenvalues are arranged in descending order

$$\lambda_1 \geq \ldots \geq \lambda_{r_g}. \tag{5.8}$$

If the true covariance matrix $\Sigma_{g(\hat{\pi})}$ is singular, some of its eigenvalues are zero. Since the true covariance matrix is unknown, a decision has to be reached on the basis

of the estimated eigenvalues. A straightforward idea is to test the null hypothesis $H_0 : \lambda_{r_g} = \lambda_{r_g-1} = \ldots = \lambda_{r_c+1} = 0$. This requires knowledge of the asymptotic distribution of the estimated eigenvalues $\hat{\lambda}_j,\ j = 1, \ldots, r_g$. Unfortunately, the asymptotic distribution of $\hat{\lambda}_j$ is known only under the assumption that all r_g eigenvalues in (5.8) take on different values (see Mardia et al. (1974)).

Alternatively, the following selection rule could be used: the rank of $\Sigma_{g(\hat{\pi})}$ equals the number of eigenvalues which are greater than or equal to some threshold value c. In other words, all eigenvalues which are smaller than some threshold value c are set to zero. Under the assumption that the estimates of the true eigenvalues converge at a rate $T^{-\alpha}$, any value $c > 0$ which goes to zero somewhat slower than $T^{-\alpha}$ can be chosen as threshold value. The latter condition ensures that $c \to 0$ as $T \to \infty$, but that the threshold value does not decline faster than the estimates $\hat{\lambda}_j$ of the true eigenvalues. Hence, even for large sample size T, some estimates of the true zero eigenvalues may still be smaller than the threshold value and are set to zero.

Extended Granger Causality

If the randomized Wald statistic is used to test the null hypothesis $H_0^{\text{EGC}} : g(\pi) = 0$ on page 65, it has to be kept in mind that the first p restrictions in $g(\pi)$ are linear. Therefore, the rank of $\Sigma_{g(\hat{\pi})}$ is not less than p, and it suffices to compare only estimates $\hat{\lambda}_j$ for $j > p$ to the threshold value c. In other words, the threshold value c should be chosen smaller than $\hat{\lambda}_p$.

Proposition 5.2: (Lütkepohl & Burda (1997, Proposition 2)).
Let $\hat{\Sigma}_{g(\hat{\pi})} = (I_{\tilde{h}} \otimes R)\hat{\Sigma}_{\hat{\pi}}(\tilde{h})(I_{\tilde{h}} \otimes R')$ be a consistent estimator of $\Sigma_{g(\hat{\pi})}$ with eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_{r_g}$ where $r_g = \tilde{h}p$ and $\tilde{h} = pk_3 + 1$. Furthermore, let $\hat{V}$ be an orthogonal matrix such that $\hat{\Sigma}_{g(\hat{\pi})} = \hat{V}\hat{\Lambda}\hat{V}'$, where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{r_g})$.
For some threshold value c with $0 < c < \hat{\lambda}_p$, define $\hat{r}_c$ to be the number of $\hat{\lambda}_j > c$ and let

$$\hat{\Lambda}_c \ = \ \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{r}_c}, 0, \ldots, 0).$$

Furthermore, let

$$\hat{\Lambda}_c^+ \;=\; \text{diag}(\hat{\lambda}_1^{-1}, \ldots, \hat{\lambda}_{\hat{r}_c}^{-1}, 0, \ldots, 0),$$

then

$$\mathcal{W}_{\text{EGC}}^+(\tilde{h}) \;=\; Tg(\hat{\pi})'\hat{V}\hat{\Lambda}_c^+\hat{V}'g(\hat{\pi}) \xrightarrow{d} \chi^2(r_c) \tag{5.9}$$

under H_0^{EGC}, where r_c is the number of eigenvalues of $\Sigma_{g(\hat{\pi})}$ greater than c.

Proof: Let $\lambda_1 \geq \ldots \geq \lambda_{r_g}$ be the eigenvalues of $\Sigma_{g(\hat{\pi})} = (I_{\tilde{h}} \otimes R)\Sigma_{\hat{\pi}}(\tilde{h})(I_{\tilde{h}} \otimes R')$ and let V be an orthogonal matrix such that $\Sigma_{g(\hat{\pi})} = V\Lambda V'$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{r_g})$. Let $\Lambda_c = \text{diag}(\lambda_1, \ldots, \lambda_{r_c}, 0, \ldots, 0)$ and let $\Lambda_c^+ = \text{diag}(\lambda_1^{-1}, \ldots, \lambda_{r_c}^{-1}, 0, \ldots, 0)$. Consistency of $\hat{\Sigma}_{g(\hat{\pi})}$ implies $\text{plim}\hat{V} = V$, $\text{plim}\hat{\Lambda}_c = \Lambda_c$ and hence $\text{plim}\hat{\Lambda}_c^+ = \Lambda_c^+$. It follows that

$$\text{plim}\hat{V}\hat{\Lambda}_c^+\hat{V}' = V\Lambda_c^+V'. \tag{5.10}$$

Denoting the first $(r_g \times r_c)$ submatrix of V by V_c with corresponding estimator $\hat{V}_c$, it follows from (5.1) that

$$\sqrt{T}V_c'\,(g(\hat{\pi}) - g(\pi)) \;\xrightarrow{d}\; N(0, \text{diag}(\lambda_1, \ldots, \lambda_{r_c})).$$

Hence, under H_0^{EGC},

$$\begin{aligned} Tg(\hat{\pi})'V_c\text{diag}(\lambda_1^{-1}, \ldots, \lambda_{r_c}^{-1})\; V_c'g(\hat{\pi}) \;&=\; Tg(\hat{\pi})'V\Lambda_c^+V'g(\hat{\pi}) \\ &\xrightarrow{d}\; \chi^2(r_c). \end{aligned} \tag{5.11}$$

Thus, the proposition follows from (5.10) and a standard limiting result for χ^2 statistics. □

Example 5.3: Trivariate VAR(1) Model.

The restrictions for y_1 being never Granger causal for y_2 have been explored in Example 5.1: there are $\tilde{h}p = 2$ restrictions such that the covariance matrix $\Sigma_{g(\hat{\pi})}$ is a real symmetric (2×2) matrix. There exists a spectral decomposition $\Sigma_{g(\hat{\pi})} = V\Lambda V'$ with $\Lambda = \text{diag}(\lambda_1,\ \lambda_2)$ and $V = [v_1, v_2]$ a real orthogonal (2×2) matrix (see Lütkepohl (1996a, 5.2.3 (5))). Since the first restriction in $g(\hat{\pi})$ is linear, $\lambda_1 \neq 0$ and $\text{rk}(\Sigma_{g(\hat{\pi})}) \geq 1$

under H_0^{EGC}.

Let $\hat{\Sigma}_{g(\hat{\pi})} = \hat{V}\hat{\Lambda}\hat{V}'$ denote the consistent estimator of $\Sigma_{g(\hat{\pi})}$, and assume that a suitable threshold value c has been determined in advance. Then if $\hat{\lambda}_2 \geq c$, the true covariance matrix $\Sigma_{g(\hat{\pi})}$ is assumed to be of full rank, and the standard Wald test can be applied. If on the other hand $\hat{\lambda}_2 < c$, it is assumed that $\text{rk}(\Sigma_{g(\hat{\pi})}) = 1$. In this case, a generalized inverse is computed as described in Proposition 5.2:

$$\begin{aligned}
\hat{\Sigma}^{+}_{g(\hat{\pi})} &= \hat{V}\hat{\Lambda}^{+}_{c}\hat{V}' \\
&= \begin{bmatrix} \hat{v}_{11} & \hat{v}_{12} \\ \hat{v}_{21} & \hat{v}_{22} \end{bmatrix} \begin{bmatrix} \hat{\lambda}_1^{-1} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{v}_{11} & \hat{v}_{21} \\ \hat{v}_{12} & \hat{v}_{22} \end{bmatrix} \\
&= \begin{bmatrix} \hat{v}_{11}\hat{\lambda}_1^{-1}\hat{v}_{11} & \hat{v}_{11}\hat{\lambda}_1^{-1}\hat{v}_{21} \\ \hat{v}_{21}\hat{\lambda}_1^{-1}\hat{v}_{11} & \hat{v}_{21}\hat{\lambda}_1^{-1}\hat{v}_{21} \end{bmatrix}.
\end{aligned}$$

Note, that only the estimate of λ_1 and its corresponding eigenvector, both which are attached to the first restriction in $g(\hat{\pi})$, enter into $\hat{\Sigma}^{+}_{g(\hat{\pi})}$. Hence, the variability of the second restriction in $g(\hat{\pi})$ is not taken into account. Instead, this restriction is only weighted with the "importance" of the first restriction in the modified Wald statistic $\mathcal{W}^{+}_{\text{EGC}}$ in (5.9).

How to Choose the Threshold Value c

A problem left open so far in the computation of a generalized covariance matrix $\Sigma^{+}_{g(\hat{\pi})}$ is to determine a suitable threshold value c. Let $\lambda_1 \geq \ldots \geq \lambda_{r_g}$ denote the eigenvalues of the $(r_g \times r_g)$ covariance matrix $\Sigma_{g(\hat{\pi})}$. Let $\hat{\Sigma}_{g(\hat{\pi})}$ be a consistent estimator of $\Sigma_{g(\hat{\pi})}$ with eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_{r_g}$, then it will be assumed in the following that

$$(\hat{\lambda}_i - \lambda_i) = O_p(T^{-1/2}) \quad \forall\, i = 1, \ldots, r_g. \tag{5.12}$$

Assumption (5.12) states that if an eigenvalue of $\Sigma_{g(\hat{\pi})}$ is greater than zero, the corresponding eigenvalue of $\hat{\Sigma}_{g(\hat{\pi})}$ will be bounded away from zero. If on the other hand an eigenvalue of $\Sigma_{g(\hat{\pi})}$ equals zero, then the corresponding eigenvalue of $\hat{\Sigma}_{g(\hat{\pi})}$ will converge towards zero at a rate $T^{-1/2}$.

The threshold value c should converge at a slower rate than the estimates of true zero eigenvalues to guarantee that eventually all estimates of true zero eigenvalues are set to zero. Hence, any threshold value $c = O_p(T^{-\epsilon})$ with $0 < \epsilon < .5$ is admissible. The

performance of different ϵ values has been analyzed for different data generating processes in a small simulation study. Results are presented in Chapter 6.

Impulse Response Analysis

The Wald test with generalized inverse can also be used to test the null hypothesis H_0^{IR} that all responses of y_2 to a one–time, one–unit impulse in y_1 are zero. The restrictions under H_0^{IR} are given in (4.32), page 75. The estimator $g(\hat{\pi}) = (I_{\bar{h}} \otimes \bar{R})\hat{\boldsymbol{\pi}}^{(\bar{h})}$ has covariance matrix

$$\begin{aligned} \Sigma_{g(\hat{\pi})} &= (I_{\bar{h}} \otimes \bar{R}) \left((\partial \boldsymbol{\pi}^{(\bar{h})} / \partial \pi') \Sigma_{\hat{\pi}} (\partial \boldsymbol{\pi}^{(\bar{h})'} / \partial \pi) \right) (I_{\bar{h}} \otimes \bar{R}') \\ &= V \Lambda V', \end{aligned}$$

where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_{\bar{h}})$, V is an orthogonal matrix of eigenvectors, and $\lambda_1 \geq \ldots \geq \lambda_{\bar{h}}$ are the eigenvalues of $\Sigma_{g(\hat{\pi})}$. Since the first restriction in $g(\hat{\pi})$ is linear, $\text{rk}(\Sigma_{g(\hat{\pi})}) \geq 1$. This leads to the following corollary:

Corollary 5.2:

Let

$$\hat{\Sigma}_{g(\hat{\pi})} = (I_{\bar{h}} \otimes \bar{R}) \left(\frac{\partial \hat{\boldsymbol{\pi}}^{(\bar{h})}}{\partial \pi'} \hat{\Sigma}_{\hat{\pi}} \frac{\partial \hat{\boldsymbol{\pi}}^{(\bar{h})'}}{\partial \pi} \right) (I_{\bar{h}} \otimes \bar{R}')$$

be a consistent estimator of $\Sigma_{g(\hat{\pi})}$ with eigenvalues $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_{\bar{h}}$ and $\bar{h} = p(k-1)$. Furthermore, let $\hat{V}$ be an orthogonal matrix such that $\hat{\Sigma}_{g(\hat{\pi})} = \hat{V}\hat{\Lambda}\hat{V}'$, where $\hat{\Lambda} = \text{diag}(\hat{\lambda}_1, \ldots, \hat{\lambda}_{\bar{h}})$.

For some threshold value c with $0 < c < \hat{\lambda}_1$, define $\hat{r}_c$ to be the number of $\hat{\lambda}_j > c$. Then a generalized inverse

$$\hat{\Sigma}^+_{g(\hat{\pi})} = \hat{V}\hat{\Lambda}^+_c\hat{V}'$$

can be computed as described in Proposition 5.2, and the Wald statistic

$$\mathcal{W}^+_{\text{IR}} = T g(\hat{\pi})' \hat{V} \hat{\Lambda}^+_c \hat{V}' g(\hat{\pi}) \xrightarrow{d} \chi^2(r_c) \tag{5.13}$$

under H_0^{IR} where r_c is the number of eigenvalues of $\Sigma_{g(\hat{\pi})}$ greater than c.

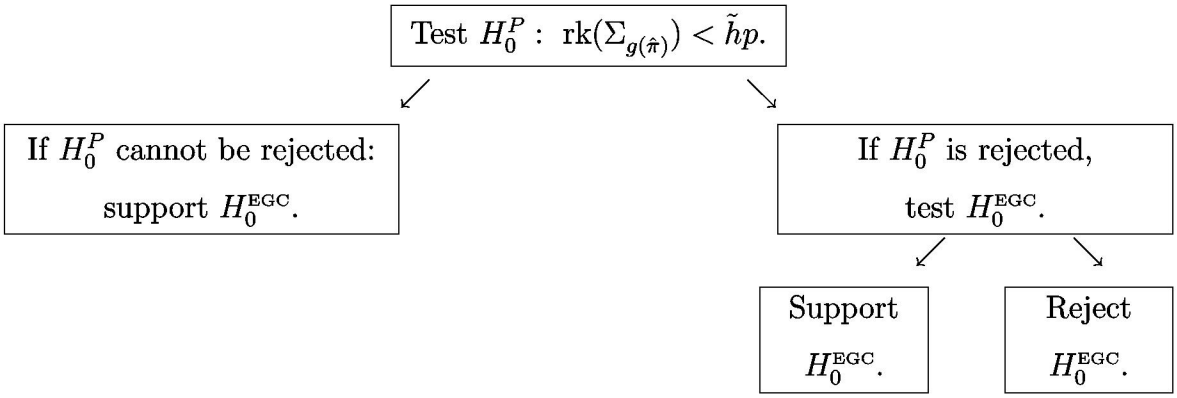


Figure 5.1: An Alternative Wald Test with Generalized Inverse.

The Wald Test with Generalized Inverse under Modification[2]

If we assume that a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ can occur only under the two conditions that there are third variables in the VAR system ($k_3 \geq 1$) and that the null hypothesis H_0^{EGC} is true, a generalized inverse may not be needed.[3] Under this assumption, a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ is a sufficient but not a necessary condition for y_1 being never Granger causal for y_2 (cf. Example 4.1). However, if a sufficient condition cannot be rejected, the corresponding null hypothesis cannot be rejected. This leads to the following strategy: in the first step, the selection rule is used to determine whether the smallest eigenvalue λ_{r_g} of $\Sigma_{g(\hat{\pi})}$ is zero. Let $H_0^P : \lambda_{r_g} = 0$ denote this hypothesis: if H_0^P cannot be rejected, the null hypothesis H_0^{EGC} is supported and the procedure terminates.

If, however, the selection rule concludes that $\Sigma_{g(\hat{\pi})}$ is regular, $\Sigma_{g(\hat{\pi})}^{-1}$ exists and the standard Wald statistic (4.11) on page 66 has a limiting $\chi^2(\tilde{hp})$–distribution. In this case, the overall null hypothesis H_0^{EGC} can be tested in the second step with a standard Wald test. Figure 5.1 illustrates this procedure.

Note, that if the selection rule is used to set up a Wald test with generalized inverse

[2]This approach follows a comment from Jean–Pierre Florens which has been gratefully acknowlegded.

[3]The assumption is based on Proposition 4.1 and Corollary 4.1, which have been derived under the null hypothesis that y_1 is never Granger causal for y_2. Note, that this assumption excludes the possibility of a singular covariance matrix $\Sigma_{g(\hat{\pi})}$ under the alternative hypothesis H_1. This, however, has not been proven.

$\Sigma^{+}_{g(\hat{\pi})}$, it is expected to deliver information on the number of zero eigenvalues. In contrast, for the present procedure the selection rule only needs to answer the question whether the smallest eigenvalue λ_{r_g} of $\Sigma_{g(\hat{\pi})}$ equals zero. Hence, the latter procedure needs less information, at least if $\tilde{h}p > 2$, than the Wald test with generalized inverse. As consequence, the risk that the Wald test in step two is based on an incorrect decision made in step one is lower. The question whether this strategy is a serious alternative to the generalized Wald statistic will be answered in the light of the simulation results in Chapter 6.

5.1.3 Other Alternative Testing Strategies

The randomized Wald test and the Wald test with generalized inverse have been developed under the premise to test jointly the set of restrictions under H_0^{EGC} with a test of size $\alpha 100\%$. However, one may think of a bulk of other strategies to test the null hypothesis that y_1 is never causal for y_2. Some alternative testing strategies are listed below, often implying a sequential testing procedure:

Alternative I: A Two–Step Procedure for Trivariate VAR(p) Models

Chapter 4 has illustrated that the problem of a possibly nonstandard asymptotic distribution of the standard Wald statistic hinges on the nonlinearity of the restrictions, at least in stationary, stable VAR models. However, if the vector of third variables is univariate, i.e. $k_3 = 1$, Corollary 3.4 presents two sets of linear restrictions ($\pi_{21,i} = \pi_{23,i} = 0$ for all $i = 1, \ldots, p$; $\pi_{21,i} = \pi_{31,i} = 0$ for all $i = 1, \ldots, p$) which characterize Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$. These restrictions can be tested subsequently with a variety of different sequential testing strategies. For instance, Giles (2000) proposes two sequential testing strategies given in Figures 5.2 and 5.3.

In Figure 5.2, the restrictions under H_{01} are tested with a standard Wald test in the first step. The corresponding Wald statistic has a limiting $\chi^2(2p)$–distribution under this null hypothesis. If H_{01} cannot be rejected, the overall null hypothesis H_0^{EGC} cannot be rejected and we conclude that y_1 is never Granger causal for y_2.
If H_{01} is rejected in the first step, the restrictions under the null hypothesis H_{02} are

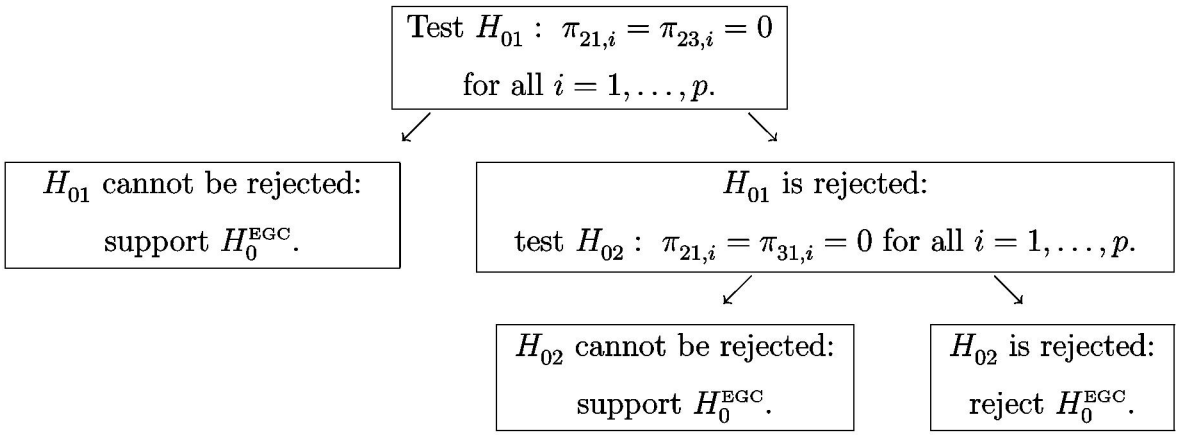


Figure 5.2: Strategy M1 (Giles (2000)).

tested in the second step. Again, the corresponding Wald statistic has a limiting $\chi^2(2p)$–distribution. If H_{02} is rejected, we conclude that y_1 is Granger causal for y_2 at some forecast horizon $h \geq 1$. Otherwise, y_1 is assumed to be never Granger causal for y_2.

Note, that H_{01} and H_{02} are not statistically independent but positively correlated: the conditional probability to reject H_{02}, given that H_{01} is false, exceeds the unconditional probability (see e.g. Krämer & Sonnberger (1986, p. 148)). Assuming an $\alpha 100\%$–significance level in both steps, the overall size of the test is less than or equal to $2\alpha 100\%$ (see Giles (2000)).

Of course, the overall null hypothesis that y_1 never Granger causes y_2 can also be tested in reversed order, i.e. by testing H_{02} in the first step and H_{01} in the second step. In practice, arranging the test sequences in a different order will affect the outcome of the test.

In contrast to Strategy M1, Strategy M2 depicted in Figure 5.3 informs about the nature of causality (direct or only indirect causality) at the expense of testing for standard Granger noncausality (H_{03}) in an extra step. Note, that the null hypotheses H_{04} and H_{05} are tested in a restricted VAR model where the coefficients $\pi_{21,i}$ are restricted to zero. This leads to lower degrees of freedom and hence may result in higher power. All corresponding Wald statistics have a limiting $\chi^2(p)$–distribution under either null hypothesis H_{03}, H_{04} or H_{05}. Assuming an $\alpha_1 100\%$–significance level for a test of H_{03} and an $\alpha_2 100\%$–significance level for a test of H_{04} and H_{05}, the overall size is bounded

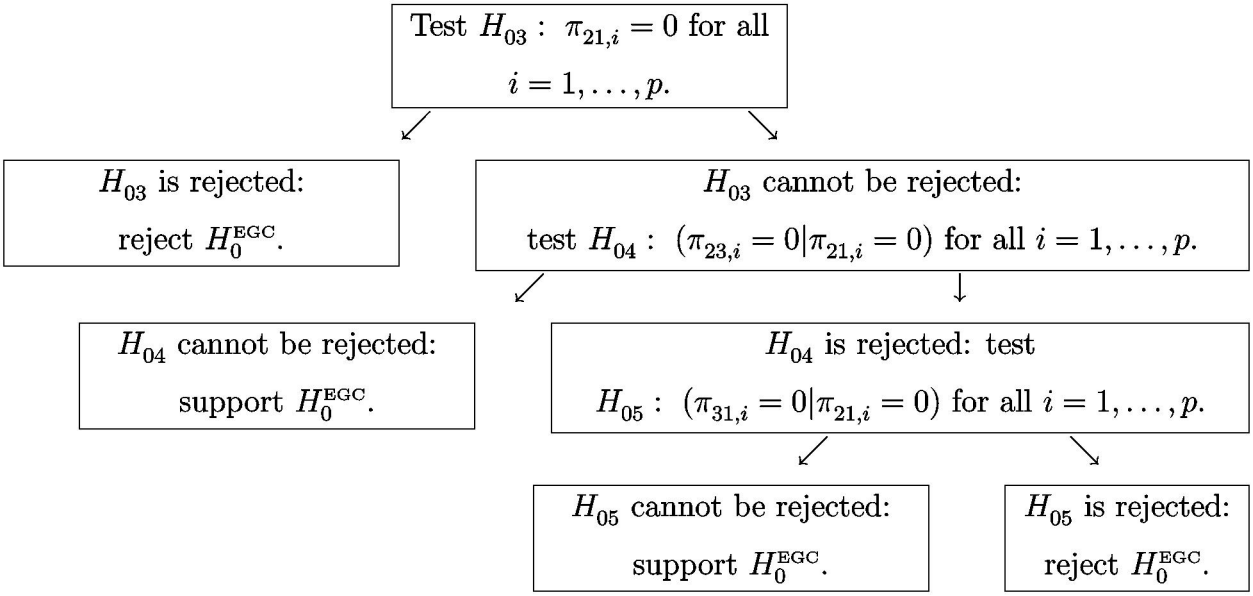


Figure 5.3: Strategy M2 (Giles (2000)).

by $\alpha = \alpha_1 + 2\alpha_2$ (see also Boudjellaba et al. (1992b)).

Both sequential testing procedures lead to conservative tests which may have empirical size much lower than the upper theoretical bound for some data generating processes (see Giles (2000), Dufour (1989), Dufour (1990) for a general discussion on bounds tests).

Alternative II: Subset VAR Models

The randomized Wald test and the Wald test with generalized inverse have been developed under the premise of testing jointly a set of necessary and sufficient restrictions. This has the advantage of avoiding the problem of unknown overall significance level inherent to sequential testing procedures. However, one may regard the selection rule used to determine the rank of $\Sigma_{g(\hat{\pi})}$ as a sequence of tests of the null hypothesis $H_0 : \lambda_j = 0$ against $H_1 : \lambda_j > 0$ for $j = \tilde{h}p,\ \tilde{h}p - 1, \ldots, p + 1$. From this point of view, the Wald test with generalized inverse is based on a sequential testing procedure of unknown overall size.

In applied work, testing for causality will be the final step in a number of steps to find a suitable model. Common tests are e.g. tests to determine the lag length p, tests

for integration or cointegration of the variables involved, and tests on whiteness and normality of residuals. The final model in which causality is tested may therefore be regarded as the result of a number of pretests. From this point of view, one may argue that adding even more tests does not matter as the overall significance level of the statistical analysis will be unknown anyway.

Moreover, VAR models require to estimate a large number of parameters. In particular in high–dimensional, high–order VAR models, the heavy parameterization may result in a low power of tests.

These arguments support Subset VAR models which allow single coefficients in the vector autoregressive coefficient matrices $\Pi_i, i = 1, \ldots, p$, to equal zero.[4] In a completely specified Subset VAR model, significance tests on all single coefficients have been performed at the model specification stage. One may therefore argue that a test of the set of restrictions under H_0^{EGC} is not necessary. Instead, it suffices to compute the vector $g(\hat{\pi})$ based on the parameter vector $\hat{\pi}$ which has been determined in the specification stage, and then to compare $g(\hat{\pi})$ to its theoretical counterpart $g(\pi) = 0$ under the null hypothesis. Under this procedure, y_1 is never Granger causal for y_2 if $g(\hat{\pi}) = 0$. An application can be found in Hsiao (1979)[5].

Application of this strategy bears the risk that noncausality of y_1 for y_2 is only a property of the Subset VAR model but need not hold in the underlying true model: in particular if a Subset VAR model is derived from a heavily parameterized VAR model, the first exclusion restrictions may be easily accepted. Hence, the order in which restrictions are tested may influence the outcome. Different specification strategies have been proposed in the literature (see e.g. Lütkepohl (1991, Chapter 5), Penm & Terrell (1982) and Penm & Terrell (1984)). Since different strategies may lead to different Subset VAR models, the outcome of such a "causality test" will depend on the strategy and will have unknown size and power properties in general (see Granger et al. (1995)).

---

[4] See e.g. Lütkepohl (1991, p. 179-189) on the specification of Subset VAR models.

[5] Hsiao (1979) uses Akaike's final prediction error (FPE) criterion to fit a vector autoregressive subset VAR model to the three time series of Canadian nominal GNP, the M2 monetary aggregate and the bank rate. Based on zero/nonzero coefficients in the subset VAR representation he then concludes to direct/indirect/no Granger causality between the three variables.

Alternative III: Bootstrapping

Bootstrapping has received widespread attention since the seminal paper of Efron (1979).[6] The bootstrap procedure can be used for instance to approximate a distribution function and to generate critical values for a test statistic (see Hinkley (1989)). This raises the question whether the bootstrap procedure can be used to approximate the limiting distribution of the standard Wald statistic in those cases where the regularity condition of the Wald test is violated?

Benkwitz et al. (2000) have shown in a related context of impulse response analysis that the usual bootstrap procedures fail if the asymptotic distribution of the impulse responses change in a discontinuous manner. Since noncausality in terms of impulse response analysis implies similar nonlinear restrictions as Granger noncausality at all forecast horizons $h \geq 1$, it is likely that the standard bootstrap procedure does not work in the present case either. A small simulation study supports this reasoning (see Appendix B).

Benkwitz et al. (2000) also show that a bootstrap procedure based on subsamples may overcome the problems of the standard bootstrap procedure, at least if the sample size is sufficiently large. The latter procedure might possibly be used for a test that y_1 is never Granger causal for y_2 under suitable modifications. It can certainly be used in the trivariate VAR(1) case where Granger noncausality at all forecast horizons coincides with noncausality in terms of impulse response analysis.

Alternative IV: Testing in a Bivariate VAR(p) Model.

Assume that y_1 indirectly Granger causes y_2 through some third variables y_3. One might wonder whether this indirect causality shows up as standard (direct) Granger causality of y_1 for y_2 if a bivariate VAR(p) model is fitted to the two variables of interest $y_{1,t}$, $y_{2,t}$? If so, this would be an easy way to get around the problem of nonlinear restrictions at the expense of loosing the information whether y_1 is directly or only indirectly Granger causal for y_2.

To answer this question, assume that the k–dimensional vector $y_t = [y_{1,t}, y_{2,t}, y'_{3,t}]'$ has

[6] See also Efron & Tibshirani (1993), Hall (1990), Hall (1992), Mammen (1992), Jeong & Maddala (1993) for an overview.

been generated by the stationary, stable VAR(p) model in (2.3), page 8, and that the matrix polynomial $\Pi(L)$ allows the following partition:

$$\Pi(L) = \begin{bmatrix} \pi_{11}(L) & \pi_{12}(L) & \pi_{13}(L) \\ \pi_{21}(L) & \pi_{22}(L) & \pi_{23}(L) \\ \pi_{31}(L) & \pi_{32}(L) & \pi_{33}(L) \end{bmatrix}. \tag{5.14}$$

Consider now the subvector $\tilde{y}_t = [y_{1,t}, y_{2,t}]$. This vector admits the ARMA($\tilde{p}, \tilde{q}$) representation

$$\tilde{\Pi}(L)\tilde{y}_t = \tilde{u}_t, \tag{5.15}$$

where

$$\begin{aligned} \tilde{\Pi}(L) &= \begin{bmatrix} \tilde{\pi}_{11}(L) & \tilde{\pi}_{12}(L) \\ \tilde{\pi}_{21}(L) & \tilde{\pi}_{22}(L) \end{bmatrix} \\ &= \begin{bmatrix} \pi_{11}(L) - \pi_{13}(L)\pi_{33}(L)^{-1}\pi_{31}(L) & \pi_{12}(L) - \pi_{13}(L)\pi_{33}(L)^{-1}\pi_{32}(L) \\ \pi_{21}(L) - \pi_{23}(L)\pi_{33}(L)^{-1}\pi_{31}(L) & \pi_{22}(L) - \pi_{23}(L)\pi_{33}(L)^{-1}\pi_{32}(L) \end{bmatrix}, \end{aligned}$$

and

$$\tilde{u}_t = \begin{bmatrix} \tilde{u}_{1,t} \\ \tilde{u}_{2,t} \end{bmatrix} = \begin{bmatrix} u_{1,t} - \pi_{13}(L)\pi_{33}(L)^{-1}u_{3,t} \\ u_{2,t} - \pi_{23}(L)\pi_{33}(L)^{-1}u_{3,t} \end{bmatrix}.$$

Three lessons can be learned:

First, fitting a bivariate VAR model to $\tilde{y}_t$ results in a misspecified model since it neglects the moving average structure of the residual vector $\tilde{u}_t$ (see Braun & Mittnik (1993)).

Second, fitting an ARMA model to $\tilde{y}_t$ and studying Granger noncausality at forecast horizon $h = 1$ therein results in nonlinear restrictions (see Granger (1969), Eberts & Steece (1984), Boudjellaba et al. (1992a, b), Lütkepohl (1994a)) and hence leads us back to the problems we sought to solve.

Third, a necessary restriction for Granger noncausality of y_1 for y_2 at forecast horizon $h = 1$ is that $\tilde{\pi}_{21}(L) \equiv 0$, where $\equiv$ means that all coefficients of the corresponding powers of L equal zero. However, since $\tilde{\pi}_{21}(L) = \pi_{21}(L) - \pi_{23}(L)\pi_{33}(L)^{-1}\pi_{31}(L)$, standard Granger noncausality in the autoregressive part of model (5.15) leaves open the possibility that y_1 is directly and indirectly Granger causal for y_2 in the true VAR(p) model but that the causal influences add up to zero. However, if $\tilde{\pi}_{21}(L) \not\equiv 0$, y_1 is directly and/or indirectly Granger causal for y_2.

5.2 Nonstationary VAR Models

It has been shown in Section 4.2.1, that the covariance matrix $\Sigma_{\hat{\pi}}$ of the least squares VAR coefficient estimates $\hat{\pi}$ is a singular matrix if the VAR model is nonstationary. As consequence, the covariance matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi')\Sigma_{\hat{\pi}}(\partial g(\pi)'/\partial \pi)$ of the set of nonlinear restrictions $g(\hat{\pi})$ may be singular even if the Jacobian matrix $(\partial g(\pi)/\partial \pi')$ is of full row rank. However, the suggestion of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) to overfit the true VAR order by the order of integration of the vector y_t can be used to get around the problem of a singular covariance matrix $\Sigma_{\hat{\pi}}$. This strategy has already been presented in Section 4.2.2, page 87, and illustrated in Example 4.2. It leads to the following testing procedure:

Under the assumption that the vector y_t is at most integrated of order one and that the true VAR lag order is p, a VAR($p+1$) model is fitted to the data in the first step. A test of the null hypothesis that y_1 is never Granger causal for y_2 is then performed in the second step using the randomized Wald test or the Wald test with generalized inverse. When setting up either test, the restrictions for Granger noncausality at all forecast horizons are imposed only on the first p coefficient matrices, hence the coefficients in $\hat{\Pi}_{p+1}$ are ignored. Moreover, the regressors belonging to the $(p+1)$–th coefficient matrix are neglected in the computation of any moment matrix, i.e. in the computation $\hat{\Sigma}_{\hat{\pi}}$.

5.3 Discussion

In this chapter, alternative testing strategies have been presented which have a known limiting distribution also in those cases where the asymptotic distribution of the standard Wald statistic is unknown. Some of these alternative testing strategies require sequential tests. Such a procedure suffers from the drawback that its overall nominal size cannot be controlled but at most bounded. Indeed, the overall nominal size may be much smaller than the theoretical size. However, a test with unknown nominal size offers not much improvement relative to the standard Wald statistic.

In contrast, the randomized Wald test and the Wald test with generalized inverse share the idea of the standard Wald test to test jointly the restrictions for Granger noncausality of y_1 for y_2 at all forecast horizons $h \geq 1$. Moreover, they have correct asymptotic

size even in those cases where the standard Wald statistic has an unknown asymptotic distribution. They should therefore be preferred to the standard Wald test.

In practice, sample sizes are finite and the exact size may differ from the theoretical size of a test. Of course, a potential user would prefer the modified Wald test which has exact size closest to the theoretical size, and largest power. Therefore, a small simulation study has been undertaken to gain insight into the small sample size and power properties of the two modified Wald tests relative to the standard Wald test. The simulation study also provides information on the sensitivity of the randomized Wald test (the Wald test with generalized inverse) towards different values of ξ (ϵ). The description of the set up, results and discussion are given in the next chapter.

Note, that the alternative testing procedures I to IV have not been included into the simulation study because it lacks the basis for comparing their size and power properties with those of the standard and modified Wald tests. Some simulation results for sequential tests of Granger noncausality in trivariate VAR models can be found in Toda & Phillips (1994) and Giles (2000).

Chapter 6

A Small Simulation Study.

6.1 Some Theoretical Considerations.

In Chapter 5, two modified Wald tests, a randomized Wald test and a Wald test with generalized inverse, have been presented: the corresponding Wald statistics $\mathcal{W}^{(\xi)}_{\mathrm{EGC}}(\tilde{h})$ and $\mathcal{W}^{+}_{\mathrm{EGC}}(\tilde{h})$ have an asymptotic χ^2–distribution under the null hypothesis H_0^{EGC} that y_1 is never Granger causal for y_2, also in those cases where the limiting distribution of the standard Wald statistic is nonstandard.

This chapter presents a simulation study on the sample size and power properties of these modified Wald tests relative to the standard Wald test. To keep the analysis simple and to exclude other factors like choice of lag length, which may influence the size and power of the tests, the simulation study has been set up for a trivariate stationary, stable VAR(1) model as studied e.g. in Examples 4.1, 5.1 and 5.3:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} \pi_{11,1} & \pi_{12,1} & \pi_{13,1} \\ \pi_{21,1} & \pi_{22,1} & \pi_{23,1} \\ \pi_{31,1} & \pi_{32,1} & \pi_{33,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \tag{6.1}$$

with $u_t \sim N(0, I_3)$. However, before details of the setup and results are presented, some theoretical considerations about the size and power of the modified Wald tests are undertaken.

To study the size, data generating processes which fulfill the null hypothesis H_0^{EGC} have been simulated (cf. page 65). For these data generating processes, the vector of

restrictions is

$$g(\pi) = \begin{bmatrix} \pi_{21,1} \\ \pi^{(2)}_{21,1} \end{bmatrix} = \begin{bmatrix} \pi_{21,1} \\ (\pi_{11,1} + \pi_{22,1})\pi_{21,1} + \pi_{23,1}\pi_{31,1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The covariance matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi')\Sigma_{\hat{\pi}}(\partial g(\pi)'/\partial \pi)$ of the estimated vector $g(\hat{\pi})$ has full rank for data generating processes where (i) either $\pi_{31,1} \neq 0$ or $\pi_{23,1} \neq 0$ and is singular for data generating processes where (ii) $\pi_{31,1} = \pi_{23,1} = 0$ holds (cf. Example 4.1).

Consider first the randomized Wald statistic $\mathcal{W}^{(\xi)}_{\mathrm{EGC}}(\tilde{h})$ given in Proposition 5.1: For $\xi = 0$, the randomized Wald test reduces to the standard Wald test which has unknown size in case (ii). For $\xi > 0$, the randomized Wald test has correct asymptotic size for any ξ value. However, in small samples, the empirical size of the test may well vary for different ξ values. The simulation study shall answer the question, in how far the size of ξ influences the empirical size of the test in small and large samples, and shall help to determine suitable ξ values.

The empirical size of the Wald test with generalized inverse (see Proposition 5.2) will depend on the performance of the selection rule which determines the number of nonzero eigenvalues and thus the rank of $\Sigma_{g(\hat{\pi})}$: Let $\lambda_1,\ \lambda_2$ denote the eigenvalues of $\Sigma_{g(\hat{\pi})}$ with $\lambda_1 > \lambda_2$. Since the first restriction in $g(\pi)$ is linear, $\partial g(\pi)/\partial \pi'$ has row rank at most equal to $p = 1$. Moreover, since $\Sigma_{\hat{\pi}}$ is nonsingular, $1 \leq \mathrm{rk}(\Sigma_{g(\hat{\pi})}) \leq 2$. Hence, $\lambda_1 \neq 0$ but $\lambda_2 \geq 0$.

The selection rule now decides on the basis of the estimated eigenvalues $\hat{\lambda}_1,\ \hat{\lambda}_2$ whether the smaller eigenvalue λ_2 equals zero (case (ii)) or not (case (i)): λ_2 is assumed to be zero whenever the estimated eigenvalue $\hat{\lambda}_2 < c$ where c is some threshold value $c = O(T^{-\epsilon})$ with $\epsilon > 0$. For a given sample size T, the size of ϵ determines the size of the threshold value: the smaller ϵ, the larger the threshold value, the more likely that $\hat{\lambda}_2 < c$ and hence the higher the probability that the selection rule decides that $\mathrm{rk}(\Sigma_{g(\hat{\pi})}) = 1$. For data generating processes where $\mathrm{rk}(\Sigma_{g(\hat{\pi})}) = 2$, choosing ϵ too small increases the probability that the rank of $\Sigma_{g(\hat{\pi})}$ will be underestimated. In contrast, for data generating processes where $\mathrm{rk}(\Sigma_{g(\hat{\pi})}) = 1$, choosing ϵ too large increases the

probability that the rank of $\Sigma_{g(\hat{\pi})}$ will be overestimated.

The selection rule may decide incorrectly if ϵ is chosen either too small or too large. However, size distortions of the Wald test with generalized inverse are expected only in the latter case:
If $\text{rk}(\Sigma_{g(\hat{\pi})}) = 2$, i.e. $\lambda_2 \neq 0$, and if the selection rule decides correctly, the Wald test with generalized inverse is identical to the standard Wald test. If $\text{rk}(\Sigma_{g(\hat{\pi})}) = 2$ but the selection rule decides that $\text{rk}(\Sigma_{g(\hat{\pi})}) = 1$, more weight is given to the first restriction in the computation of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$. However, as long as the true data generating process fulfills the restrictions under H_0, this does not affect the size, because the Wald test with generalized inverse uses the number of nonzero eigenvalues as number of degrees of freedom. Hence, if λ_2 is assumed to be zero, the test compares the value of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$ to the critical value of a $\chi^2(1)$–distribution.
If the selection rule overestimates the rank of $\Sigma_{g(\hat{\pi})}$, the Wald test with generalized inverse behaves like the standard Wald test. In particular, it suffers from the same size distortions.
Different values for ϵ will be used in the simulation study to analyze the influence of this parameter on the rank decision of the selection rule and the size of the Wald test with generalized inverse.

To study the power properties of the modified Wald tests, $g(\pi)$ is chosen equal to $\delta/\sqrt{T}$ with δ a nonzero (2×1) vector, i.e.

$$H_1 : \; g(\pi) = \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}. \tag{6.2}$$

Under this alternative, the standard Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ has a noncentral $\chi^2(2, \gamma)$–distribution with

$$\gamma \;=\; \delta' \Sigma^{-1}_{g(\hat{\pi})} \delta.$$

Note, that under the alternative hypothesis (6.2), the Jacobian matrix $\partial g(\pi)/\partial \pi'$ always has full row rank. Hence, $\Sigma_{g(\hat{\pi})}$ is regular and $\Sigma^{-1}_{g(\hat{\pi})}$ exists.

The randomized Wald statistic $\mathcal{W}^{(\xi)}_{\text{EGC}}(\tilde{h})$ has a noncentral $\chi^2(2,\ \gamma^{(\xi)})$–distribution with

$$\gamma^{(\xi)} \quad = \quad \delta' \left[\Sigma_{g(\hat{\pi})} + \xi\Sigma_w\right]^{-1} \delta.$$

The basic principle of the randomized Wald statistic $\mathcal{W}^{(\xi)}_{\text{EGC}}(\tilde{h})$ consists of adding random noise to the set of restrictions $g(\hat{\pi})$ in the standard Wald statistic, thus rendering the estimation of $g(\pi)$ less efficient. Due to this inefficiency, we expect the power of the randomized Wald statistic to be smaller than the power of the standard Wald statistic. Moreover, the power should decrease as the amount of inefficiency increases which is added to $g(\hat{\pi})$, hence with increasing ξ.

For data generating processes which fulfill the alternative hypothesis H_1, the Wald statistic with generalized inverse $\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$ follows a noncentral $\chi^2(r_c, \gamma^+)$–distribution with r_c the number of nonzero eigenvalues of $\Sigma^+_{g(\hat{\pi})}$, and

$$\gamma^+ \quad = \quad \delta'\Sigma^+_{g(\hat{\pi})}\delta.$$

Under the alternative hypothesis (6.2), $\text{rk}(\Sigma_{g(\hat{\pi})}) = 2$, and the Wald test with generalized inverse behaves just as the standard Wald test whenever the selection rule estimates the rank of $\Sigma_{g(\hat{\pi})}$ correctly.

If the selection rule underestimates the rank of $\Sigma_{g(\hat{\pi})}$, the Wald test with generalized inverse follows asymptotically a noncentral $\chi^2(1, \gamma^+)$–distribution while the standard Wald test is asymptotically $\chi^2(2, \gamma)$–distributed. Since the Wald test with generalized inverse has fewer degrees of freedom, it may even have more power than the standard Wald test, depending on the noncentrality parameter γ^+. In particular, restricting the smaller eigenvalue of $\Sigma_{g(\hat{\pi})}$ to zero amounts to "weighting the vector δ only with the importance of the first restriction in violating the null hypothesis":

For data generating processes which violate the first restriction of the null hypothesis ($\delta_1 \neq 0$), putting more weight on the first restriction in the computation of the Wald statistic may raise the power of the Wald test with generalized inverse, relative to the power of the standard Wald statistic. Gallant (1977) and Gallant & Tauchen (1989) have pointed out in a related context of redundant restrictions, that the increase in power is positively correlated with the relative importance of the first restriction over the second one in violating the null hypothesis. For instance, the increase in power will

be higher if $\delta_2 = 0$ than if $\delta_2 \neq 0$.

For data generating processes where $\delta_1 = 0$ but $\delta_2 \neq 0$, the picture is reversed: since these data generating processes fulfill the first restriction under H_0^{EGC}, giving more weight to the first restriction in the computation of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$ results in a low noncentrality parameter and hence in lower power relative to the standard Wald statistic.

In general, underestimating the rank of $\Sigma_{g(\hat{\pi})}$ reduces the power of $\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$ relative to $\mathcal{W}_{\text{EGC}}(\tilde{h})$ the more, the more weight is put on restrictions which are fulfilled under the alternative.

6.2 Simulation Setup and Simulation Results.

In the simulation study[1], the following three–dimensional VAR(1) models have been used:

$$\text{Model I:} \quad \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} \pi_{11,1} & 0 & 0 \\ \pi_{21,1} & \pi_{22,1} & 0 \\ \pi_{31,1} & .5 & \pi_{33,1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \tag{6.3}$$

with $\pi_{ii,1} \in \{-.9, -.5, -.3, .3, .5, .9\}$ the eigenvalues of the process and $\pi_{31,1} \in \{.5, 0\}$, and

$$\text{Model II:} \quad \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} .4 & .4 & .3 \\ 0 & .4 & \pi_{23,1} \\ .7 & .5 & .4 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}. \tag{6.4}$$

Different values for $\pi_{21,1}$ and $\pi_{23,1}$ have been chosen to analyze the size of the modified Wald tests for data generating processes where $\Sigma_{g(\hat{\pi})}$ is regular and for those where $\Sigma_{g(\hat{\pi})}$ is singular, as well as to analyze the power of the modified Wald tests. Exact figures will be given later.

For each model, data have been generated in the following way: $T + B$ observations have been drawn for a trivariate residual series from a standard normal distribution. With the so obtained $(3(T + B) \times 1)$ vector $u = [u'_{-B+1},\ u'_{-B+2}, \ldots, u'_0, u'_1, \ldots, u'_T]'$

[1] All simulations have been carried out with Gauss 3.2 for UNIX.

and the coefficient matrix Π_1 of Model I respectively Model II, $(T + B)$ data for the trivariate series y_t have been generated according to (6.1) with starting values set to zero. The first $B = 100$ presample values have then been cut off to eliminate the starting–up effects.

In the next step, a trivariate VAR(1) model with intercept has been fitted to the generated data by ordinary least squares. Based on the least squares estimates the standard Wald statistic ($\mathcal{W}_{\text{EGC}}(\tilde{h})$), the randomized Wald statistic ($\mathcal{W}^{(\xi)}_{\text{EGC}}(\tilde{h})$) and the Wald statistic with generalized inverse ($\mathcal{W}^{+}_{\text{EGC}}(\tilde{h})$) were computed:

Computation of the randomized Wald statistic has been performed for $\xi \in \{.01, .1, .5, 10\}$ to analyze the sensitivity with respect to different ξ values. The resulting Wald statistics are denoted by $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01), \mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1), \mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$ and $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$.

Computation of the Wald test with generalized inverse has been based on the following selection rule: let $\hat{\lambda}_2$ be the smaller eigenvalue of $\hat{\Sigma}_{g(\hat{\pi})}$, then $\hat{\lambda}_2$ is restricted to zero whenever $\hat{\lambda}_2 < c$ where $c = T^{-\epsilon}$. Different values $\epsilon \in \{.33, .5, .67, .75, .9, 1\}$ have been used in the simulation.[2] The corresponding Wald statistics are denoted by $\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33), \ldots, \mathcal{W}^{+}(\epsilon = 1)$. A Wald test with $\hat{\lambda}_2$ always set to zero is also included for comparison purposes ($\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$), as well as the standard Wald statistic ($\mathcal{W}_{\text{EGC}}$).

For all different Wald statistics, p–values have been computed and compared to the theoretical size of 5%: the null hypothesis H_0^{EGC} that y_1 is never Granger causal for y_2 is rejected if the p–value is smaller than 5%. This procedure has been repeated $N = 10000$ times to calculate the relative rejection frequency of each statistic.

In each replication, pseudo standard normal random numbers have been generated for the trivariate residual series u_t. Thereby, the pseudo random number generator has been initialized to ensure that the same set of 10000 residual series has been used throughout and thus to ease the comparability of different models and scenarios. Moreover, in each replication computation of all test statistics is based on the same data set to ensure that all tests are computed under the same conditions.

Different sample sizes $T \in \{100,\ 500,\ 1000,\ 2000\}$ have been considered to study the

[2] According to assumption (5.12), any threshold value $c = O_p(T^{-\epsilon})$ with $0 < \epsilon < .5$ is admissible. In the simulation study, a wider range of ϵ values has been considered to investigate inhowfar the results support assumption (5.12).

size as well as the power of the modified Wald tests.

Simulation of Size for Model I

Consider first Model I in (6.3): since Π_1 is chosen lower triangular, the eigenvalues of Π_1 are simply the coefficients on the main diagonal. Choosing these coefficients smaller than one in absolute value ensures stationarity of the VAR process. Different eigenvalues have been considered to study processes close to the nonstationary region ($\pi_{ii,1} \in \{-.9, .9\}$), as well as processes which are well inside the stationary region ($\pi_{ii,1} \in \{-.5, -.3, .3, .5\}$). Note that in Model I, $\pi_{23,1} = 0$. Choosing $\pi_{21,1} = 0$ means that the restrictions for Granger noncausality at all forecast horizons are fulfilled. In this case, the coefficient $\pi_{31,1}$ controls whether the covariance matrix $\Sigma_{g(\hat{\pi})}$ is regular ($\pi_{31,1} = .5$) or singular ($\pi_{31,1} = 0$), see Example 4.1. The coefficient $\pi_{32,1}$ is arbitrarily set to .5. Other values for $\pi_{32,1}$ did not change the results significantly.
To study the size of the different Wald statistics, data have been generated for $\pi_{21,1} = 0$, $\pi_{31,1} \in \{0, .5\}$ and two different sample sizes $T = 100, 1000$. For $T = 1000$, all tests should have empirical size close to the nominal size of 5%. $T = 100$ has been chosen to study the empirical size of the different Wald statistics for a sample size more likely in empirical applications. Results are given in Tables 6.1 and 6.2. Note, that for 10000 independent replications the standard deviation of a 5% rejection probability is $\sqrt{.05 \times .95/10000} \times 100\% = .22\%$.

In the left–hand sides of Tables 6.1 and 6.2, $\pi_{31,1} = .5$. Therefore, all Wald tests should have empirical size well inside a 2–standard–deviation bound around 5%:
The results of Table 6.1 show that for $T = 1000$ the empirical sizes of all Wald statistics are mostly inside the 2–standard–deviation bound around 5% for any ξ value and any $\epsilon < 1$. Only the Wald test with generalized inverse shows some size distortions if the data generating process is close to the nonstationary border ($\pi_{ii,1} \in \{-.9, .9\}$). A comparison with the empirical size of $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ shows that the selection rule underestimates the rank of $\Sigma_{g(\hat{\pi})}$ for these processes, regardless of the ϵ value.

Results differ for $T = 100$: for processes close to the nonstationary border, all Wald

Table 6.1

Size of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 1000$.

	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} =$						$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} =$					
	-.9	-.5	-.3	.3	.5	.9	-.9	-.5	-.3	.3	.5	.9
$\mathcal{W}_{\text{EGC}}$	5.2	4.5	5.0	5.2	5.2	5.3	1.6	1.6	1.6	1.7	1.8	1.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	5.3	4.6	5.0	5.2	5.3	5.3	4.8	3.9	3.6	3.7	4.0	4.8
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	4.9	4.9	5.2	5.5	5.5	5.3	5.4	5.2	5.1	5.1	5.1	5.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	5.1	5.1	5.4	5.6	5.3	5.3	5.4	5.4	5.3	5.3	5.3	5.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.1	5.3	5.2	5.5	5.3	5.3	5.4	5.4	5.3	5.4	5.3	5.2
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	4.7	5.1	5.0	5.2	5.5	5.3	5.2	5.0	5.2	5.5	5.4	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	4.7	4.5	5.0	5.2	5.2	5.3	5.2	5.0	5.2	5.5	5.4	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	4.7	4.5	5.0	5.2	5.2	5.3	5.2	5.0	5.2	5.4	5.4	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	4.7	4.5	5.0	5.2	5.2	5.3	5.2	5.0	2.4	2.7	5.4	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	4.7	4.5	5.0	5.2	5.2	5.3	5.2	1.6	1.6	1.7	1.8	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	4.2	4.5	5.0	5.2	5.2	4.7	5.2	1.6	1.6	1.7	1.8	5.0
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	4.7	5.1	5.2	5.6	5.5	5.3	5.2	5.0	5.2	5.5	5.4	5.0

Table 6.2

Size of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 100$.

	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} =$						$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} =$					
	-.9	-.5	-.3	.3	.5	.9	-.9	-.5	-.3	.3	.5	.9
$\mathcal{W}_{\text{EGC}}$	6.2	4.4	3.7	4.1	4.4	8.9	3.7	2.0	2.0	1.9	2.0	5.0
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	6.1	4.8	4.0	4.2	4.7	8.5	4.9	2.3	2.1	2.1	2.2	6.0
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	6.3	5.1	4.9	4.8	5.2	8.1	6.5	4.2	3.9	3.8	4.2	8.0
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	6.4	5.4	5.3	5.3	5.6	8.2	7.0	5.2	5.0	5.2	5.3	8.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	6.5	5.6	5.5	5.5	5.8	7.9	7.1	5.6	5.3	5.6	5.6	8.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	7.8	5.8	5.3	5.7	6.1	9.6	8.6	5.8	5.8	6.0	6.1	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	7.8	5.3	3.9	4.4	5.4	9.6	8.6	5.7	5.5	5.6	6.1	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	7.8	4.5	3.7	4.1	4.6	9.6	8.6	5.3	2.8	2.5	5.3	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	7.8	4.4	3.7	4.1	4.4	9.6	8.6	3.9	2.2	1.9	3.9	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	7.8	4.4	3.7	4.1	4.4	9.5	8.6	2.1	2.0	1.9	2.1	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	7.7	4.4	3.7	4.1	4.4	9.5	8.6	2.1	2.0	1.9	2.0	10.6
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	7.8	5.8	5.6	5.8	6.1	9.6	8.6	5.8	5.8	6.0	6.1	10.6

tests now show a tendency to overreject. For processes well inside the stationary border, Table 6.2 shows that the choice of ξ and ϵ matters in small samples.

The empirical size of the randomized Wald test is now well inside the 2–standard–deviation bound around 5% only for $\xi \in \{.1, .5\}$. For ξ too big ($\xi = 10$), the randomized Wald test shows a tendency to overreject. For ξ too small ($\xi = .01$), the empirical size of the randomized Wald test tends towards the size of the standard Wald test which behaves conservatively for this sample size, at least if the process is well inside the stationary region.

The empirical size of the Wald test with generalized inverse equals that of the standard Wald test if $\epsilon \geq .5$. Hence, for these ϵ values, the rank of $\Sigma_{g(\hat{\pi})}$ seems to be estimated correctly. For $\epsilon = .33$, comparison with the results for $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ shows that the selection rule tends to underestimate the rank of $\Sigma_{g(\hat{\pi})}$. In this case, the value of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$ is compared to the smaller critical value of a $\chi^2(1)$–distribution. This leads to an empirical size which significantly exceeds the nominal size of 5%.

In the right–hand sides of Tables 6.1 and 6.2, $\pi_{31,1} = 0$ and the standard Wald statistic fails to have an asymptotic $\chi^2(2)$–distribution. This is clearly seen in Table 6.1: using a critical value from a $\chi^2(2)$–distribution leads to a standard Wald test with empirical size much less than the nominal significance level.

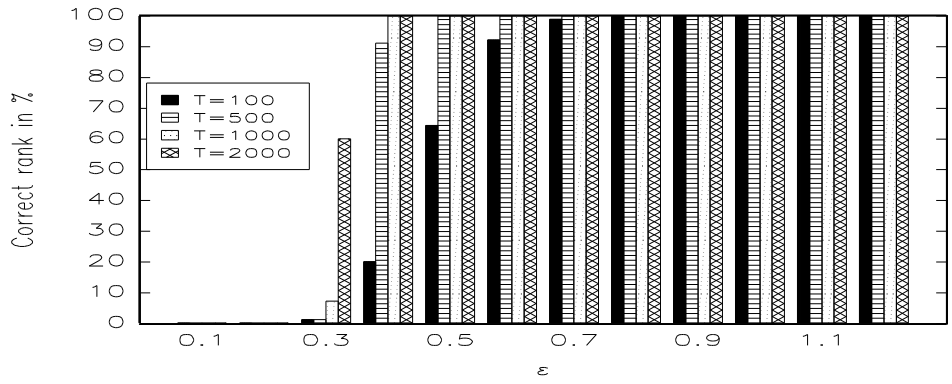
In contrast, the empirical size of the randomized Wald test improves already for a ξ value as small as .01. For $\xi \in \{.1, .5, 10\}$, the empirical size lies well inside a 2–standard–deviation bound around 5%.

Since the true rank of $\Sigma_{g(\hat{\pi})}$ equals one, the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$, which is computed under this assumption, should have empirical size close to the nominal size of 5% and should be taken as yardstick for comparison. Indeed, the empirical size of $\mathcal{W}^{+}(\text{rk} = 1)$ is always inside the 2–standard–deviation bound around 5%. The same results are obtained for $\epsilon \in \{.33, .5, .67\}$. Hence for these ϵ values, the true rank of $\Sigma_{g(\hat{\pi})}$ is estimated correctly. For larger ϵ values, the rank is overestimated. As consequence, the Wald test with generalized inverse behaves as conservatively as the standard Wald test. This holds in particular for $\epsilon = \{.9, 1.0\}$ and processes well inside the stationary

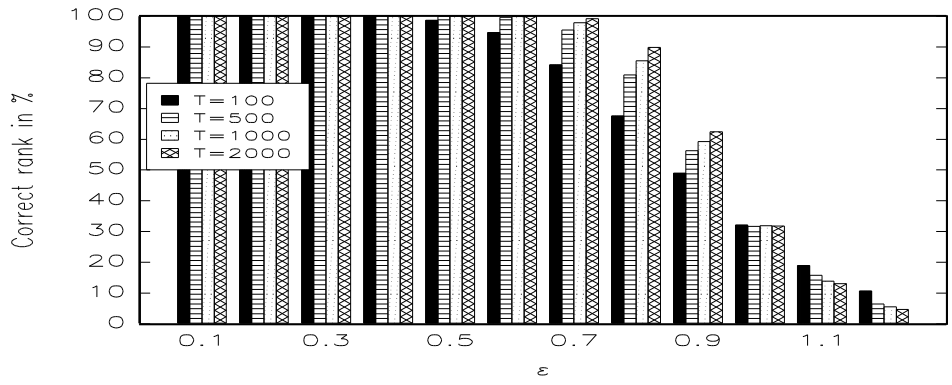
region. Note that the good empirical size for processes close to the nonstationary border is deceiving: it results from underestimation of the rank of $\Sigma_{g(\hat{\pi})}$ which turns out favorable in this case.

By and large, the same picture is also obtained for a sample size of $T = 100$:
For the randomized Wald test, a value $\xi = .01$ again proves to be too small to correct the size distortions of the standard Wald test. Choosing $\xi = .1$ results in a clear improvement in this respect, but best results are obtained for $\xi = .5$.
The Wald test with generalized inverse shows a strong tendency to overreject if the process is close to the nonstationary border. For processes well inside the stationary border, it performs best for $\epsilon \in \{.33, .5\}$. However, choosing $\epsilon \in \{.67, .75\}$ still guarantees that the Wald test with generalized inverse has empirical size closer to the nominal size of 5% than the standard Wald statistic. For $\epsilon > .75$, the Wald test with generalized inverse behaves as conservatively as the standard Wald test.

Tables 6.1 and 6.2 show that the empirical size of the Wald test with generalized inverse varies strongly for different ϵ values, in particular if T is small ($T = 100$). In practical applications, sample sizes of $T = 100$ observations are quite common, and the choice of ϵ may significantly influence the outcome. Therefore, a closer look shall be thrown on the empirical size of the Wald test with generalized inverse for $\epsilon \in \{0.1, 0.2, \ldots, 1.2\}$, different sample sizes $T \in \{100, 500, 1000, 2000\}$ and a nominal significance level of $\alpha = 5\%$. However, the size of the Wald test with generalized inverse does not give a complete picture of the performance of the selection rule: if the selection rule underestimates the true rank of $\Sigma_{g(\hat{\pi})}$, the Wald test with generalized inverse still has correct asymptotic size since it compares the value of the Wald statistic $\mathcal{W}^{+}_{\mathrm{EGC}}(\epsilon)$ to the critical value of a $\chi^2(1)$–distribution in this case. Therefore, the performance of the selection rule for different ϵ values and different sample sizes is also depicted.
Since Tables 6.1 and 6.2 draw mostly the same picture for eigenvalues $\pi_{ii,1} \in \{.3, .5, .9\}$ as for eigenvalues $\pi_{ii,1} \in \{-.3, -.5, -.9\}$, results have been obtained only for the latter range of eigenvalues. Results for Model I with eigenvalues $\pi_{ii,1} = -.3$ are given in Figures 6.1 and 6.2:



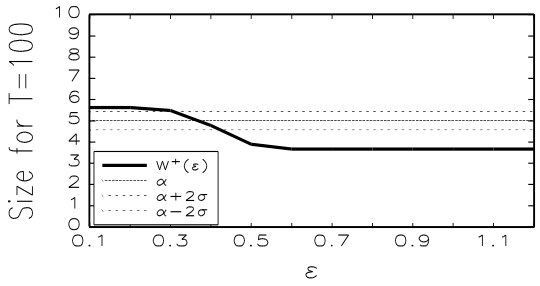
(a) $\pi_{31,1} = .5$



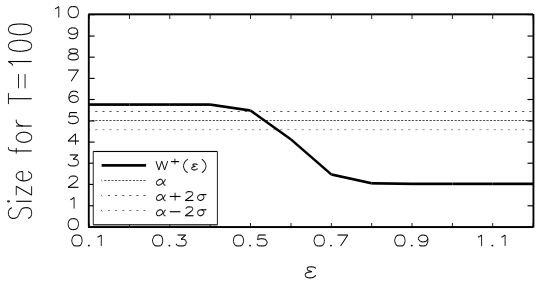
(b) $\pi_{31,1} = 0$

Figure 6.1: Performance of Selection Rule for Model I ($\pi_{ii,1} = -.3$).

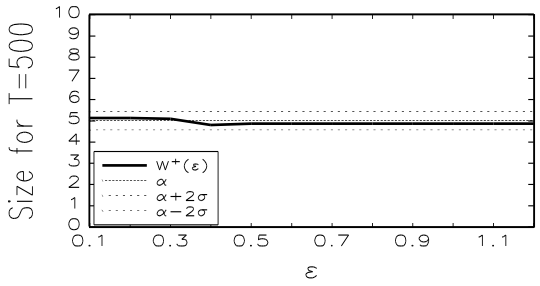
If $\pi_{31,1} = .5$, $\text{rk}(\Sigma_{g(\hat{\pi})}) = 2$. Figure 6.1 (a) shows that for sample sizes $T \geq 500$ and $\epsilon \geq .5$, the selection rule correctly estimates the rank of $\Sigma_{g(\hat{\pi})}$. If $T = 100$, the same result only holds for $\epsilon > .7$. This does not come as a surprise, given that the threshold value $c = T^{-\epsilon}$ depends on T as well as on ϵ: the larger T, the smaller ϵ can be chosen to obtain the same threshold value, and hence the same decision of the selection rule. The empirical size of the Wald test with generalized inverse should be close to the nominal size, even if the selection rule underestimates the true rank of $\Sigma_{g(\hat{\pi})}$: in the latter case, the Wald test with generalized inverse compares the value of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\epsilon)$ to the critical value of a $\chi^2(1)$–distribution. This is clearly seen in Figures 6.2 (c), (e) and (g) where the empirical size is well inside the 2–standard–deviation



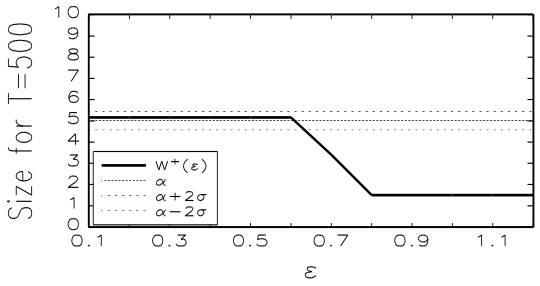
(a) $\pi_{31,1} = .5$



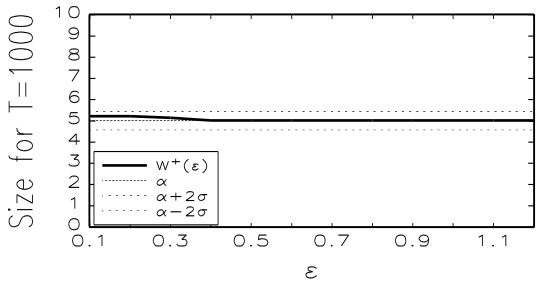
(b) $\pi_{31,1} = 0$



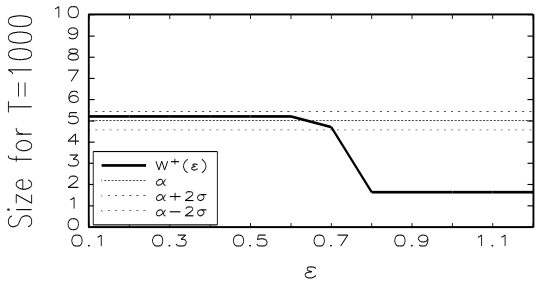
(c) $\pi_{31,1} = .5$



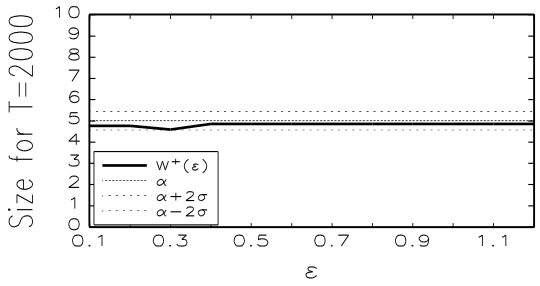
(d) $\pi_{31,1} = 0$



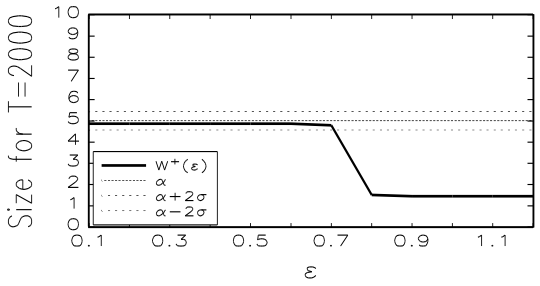
(e) $\pi_{31,1} = .5$



(f) $\pi_{31,1} = 0$



(g) $\pi_{31,1} = .5$



(h) $\pi_{31,1} = 0$

Figure 6.2: Size of $\mathcal{W}^{+}_{\text{EGC}}(\epsilon)$ in Model I ($\pi_{ii,1} = -.3$).

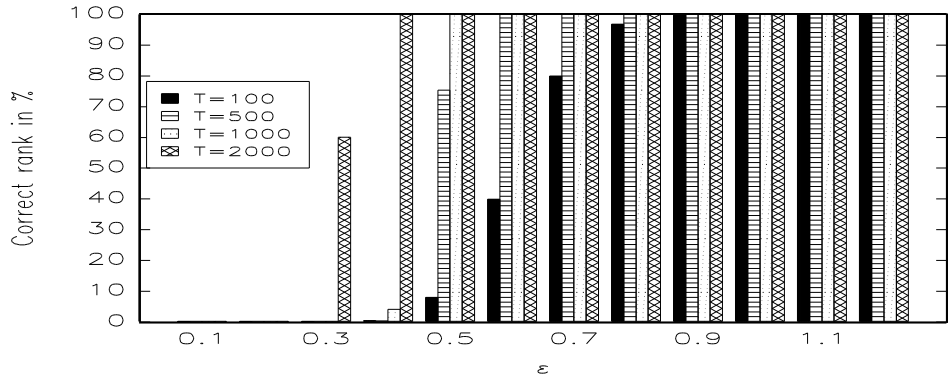
bound around 5%, even for small ϵ values. In contrast, Figure 6.2 (a) shows that the Wald statistic with generalized inverse behaves conservatively if $\epsilon \geq .4$ although the selection rule correctly recognizes the true rank of $\Sigma_{g(\hat{\pi})}$. Table 6.2 solves the puzzle: for $T = 100$, the standard Wald test behaves conservatively for the present DGP. However, the Wald test with generalized inverse is identical to the standard Wald test if $\Sigma_{g(\hat{\pi})}$ is regular and if the selection rule correctly estimates the rank of $\Sigma_{g(\hat{\pi})}$.

If $\pi_{31,1} = 0$, the true rank of $\Sigma_{g(\hat{\pi})}$ equals one. The selection rule recognizes the true rank of $\Sigma_{g(\hat{\pi})}$ for any $\epsilon \leq .6$ and sample sizes $T \geq 500$. For $T = 100$, choosing $\epsilon = .6$ still guarantees a correct rank estimation in 90% of all replications. For larger ϵ values, the threshold value c becomes too small. As consequence, the true rank of $\Sigma_{g(\hat{\pi})}$ is overestimated so that the Wald test with generalized inverse behaves just as the standard Wald test. Tables 6.1 and 6.2 show that the latter test is conservative for the present DGP. This is clearly seen in Figures 6.2 (b), (d), (f) and (h): the kinks in the size functions occur at those ϵ values which mark the incorrect rank estimation of the selection rule.

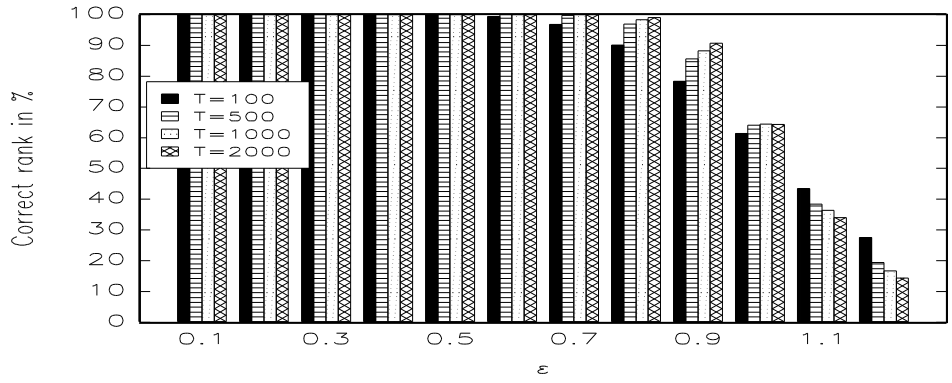
By and large, the same picture is obtained for eigenvalues $\pi_{ii,1} = -.5$ in Model I: For $\pi_{31,1} = .5$, Figure 6.3 (a) shows that the true rank of $\Sigma_{g(\hat{\pi})}$ is correctly estimated in 100% of all replications if $\epsilon \geq .6$ and $T \geq 500$. If $\epsilon = .5$ and $T = 500$, the same result only holds in 75% of all replications. If $T = 100$, ϵ has to be chosen very large to make the threshold value c sufficiently small: $\epsilon \geq .8$ guarantees a correct decision of the selection rule in at least 90% of all replications. These figures demonstrate the sensitivity of the selection rule with respect to ϵ in different samples.

However, Figures 6.4 (a), (c), (e) and (g) show that there is hardly any cost involved in choosing a "wrong" ϵ value because underestimating the rank of $\Sigma_{g(\hat{\pi})}$ only slightly affects the size of the Wald test with generalized inverse: indeed, the empirical size is almost everywhere inside the 2–standard–deviation bound around 5% for any ϵ value and sample size T.

For $\pi_{31,1} = 0$, Figure 6.3 (b) shows that the true rank is correctly estimated in at least 95% of all replications if $\epsilon \leq .7$, regardless of the sample size. For higher ϵ values, the selection rule overestimates the rank of $\Sigma_{g(\hat{\pi})}$ so that the Wald test with generalized

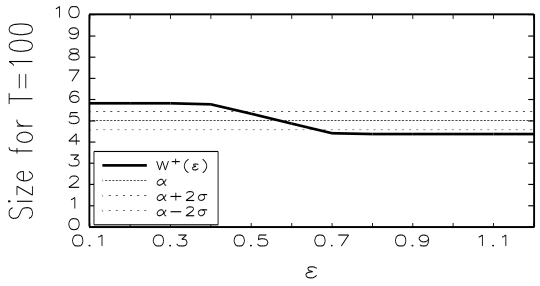


(a) $\pi_{31,1} = .5$

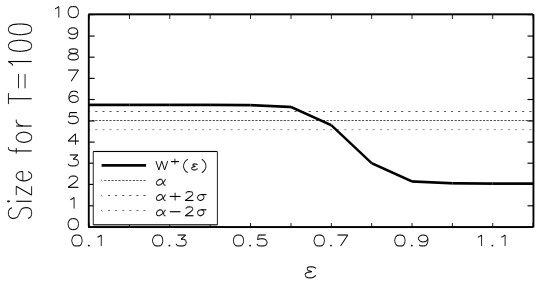


(b) $\pi_{31,1} = 0$

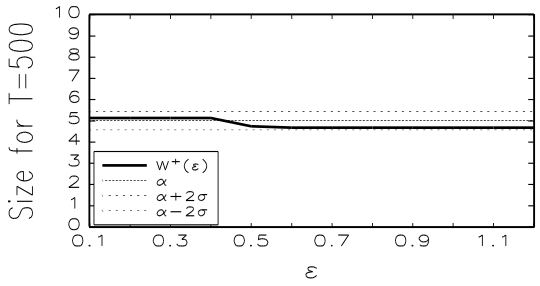
Figure 6.3: Performance of Selection Rule for Model I ($\pi_{ii,1} = -.5$).



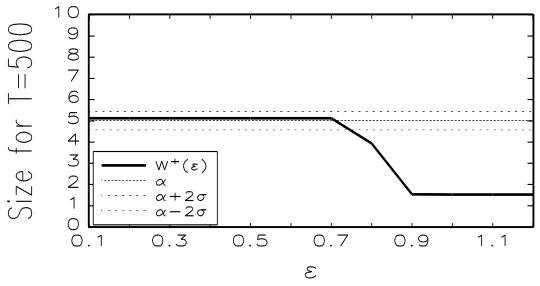
(a) $\pi_{31,1} = .5$



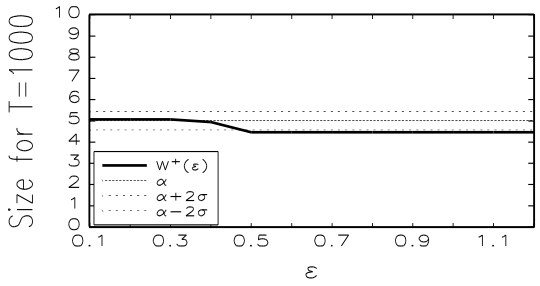
(b) $\pi_{31,1} = 0$



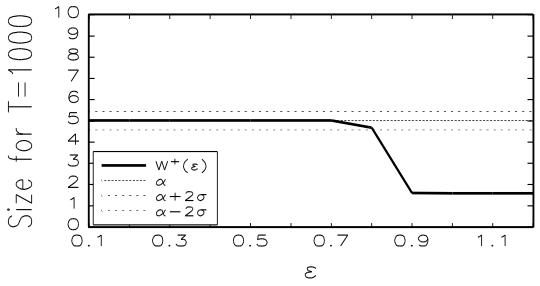
(c) $\pi_{31,1} = .5$



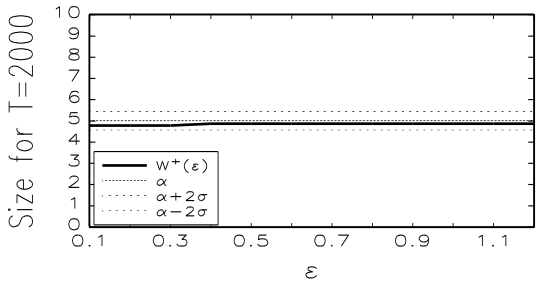
(d) $\pi_{31,1} = 0$



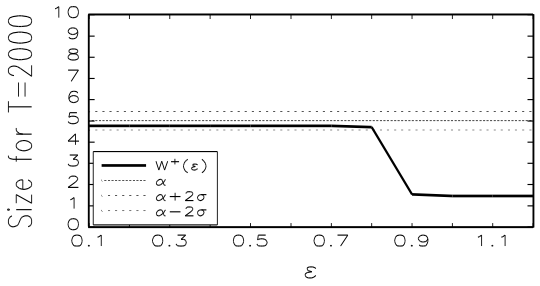
(e) $\pi_{31,1} = .5$



(f) $\pi_{31,1} = 0$

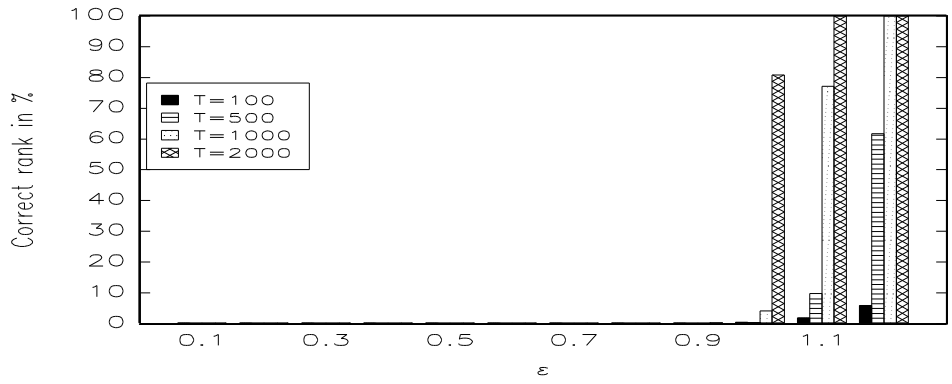


(g) $\pi_{31,1} = .5$

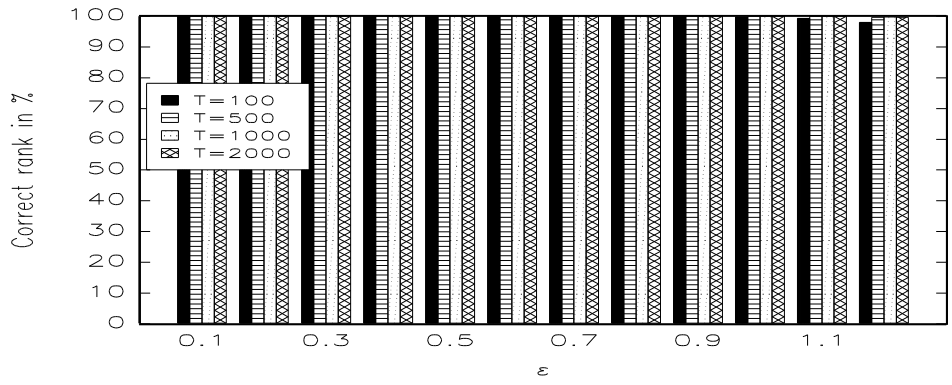


(h) $\pi_{31,1} = 0$

Figure 6.4: Size of $\mathcal{W}^{+}_{\text{EGC}}(\epsilon)$ in Model I ($\pi_{ii,1} = -.5$).



(a) $\pi_{31,1} = .5$

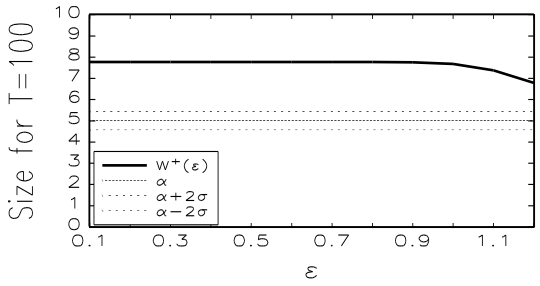


(b) $\pi_{31,1} = 0$

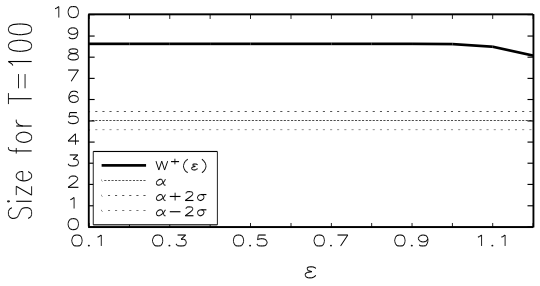
Figure 6.5: Performance of Selection Rule for Model I ($\pi_{ii,1} = -.9$).

inverse behaves just as the standard Wald test, hence is conservative. This effect shows up in a kinked size function in Figures 6.4 (b), (d), (f) and (h).

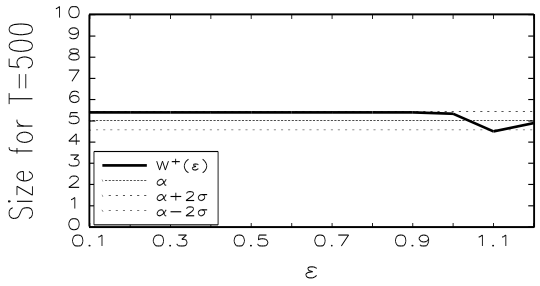
Results differ if the VAR process is close to the nonstationary border ($\pi_{ii,1} = -.9$): Figure 6.5 (a) shows that for $\pi_{31,1} = .5$ and $\epsilon \leq 1$, the rank of $\Sigma_{g(\hat{\pi})}$ is always underestimated. This result arises because the second (smaller) eigenvalue of $\Sigma_{g(\hat{\pi})}$ is close to zero even in the regular case where $\pi_{31,1} = .5$. This can be seen in Table 6.3, which shows the average estimated eigenvalue $\bar{\hat{\lambda}}_2$ for different data generating processes and different sample sizes. Thereby, $\bar{\hat{\lambda}}_2 = (1/N) \sum_{n=1}^{N} \hat{\lambda}_2^{(n)}$, with $\hat{\lambda}_2^{(n)}$ the smaller eigenvalue of $\hat{\Sigma}_{g(\hat{\pi})}$ in the n–th replication, has been computed for $N = 10000$ replications.
Table 6.3 shows that for $\pi_{ii,1} = -.9$, λ_2 is close to zero and so is $\bar{\hat{\lambda}}_2$. This explains why



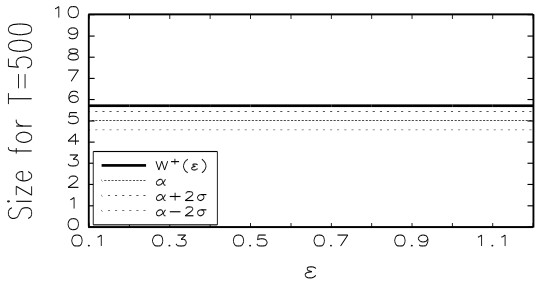
(a) $\pi_{31,1} = .5$



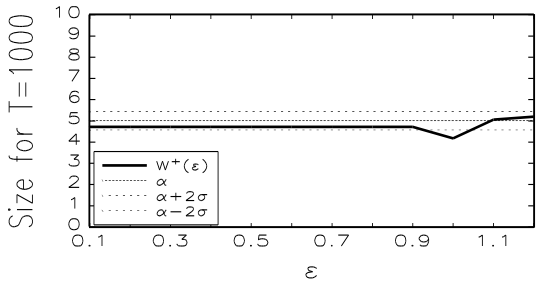
(b) $\pi_{31,1} = 0$



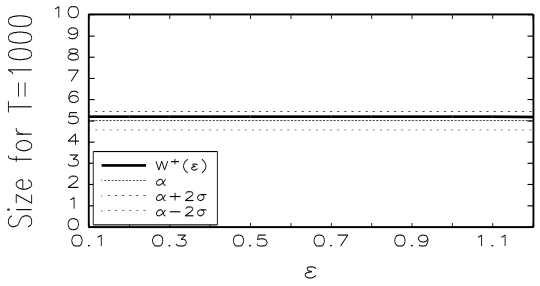
(c) $\pi_{31,1} = .5$



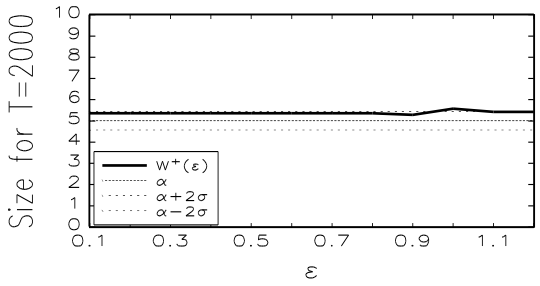
(d) $\pi_{31,1} = 0$



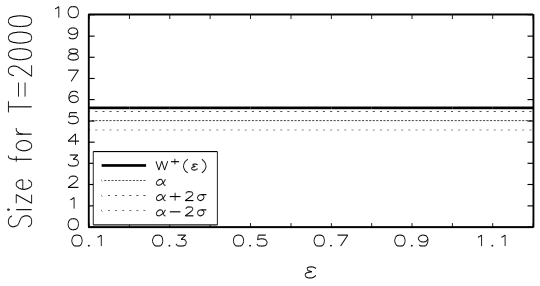
(e) $\pi_{31,1} = .5$



(f) $\pi_{31,1} = 0$



(g) $\pi_{31,1} = .5$



(h) $\pi_{31,1} = 0$

Figure 6.6: Size of $\mathcal{W}^{+}_{\text{EGC}}(\epsilon)$ in Model I ($\pi_{ii,1} = -.9$).

Table 6.3

$\bar{\hat{\lambda}}_2$ and λ_2 in Model I

$\pi_{31,1} = 0$			
	$\pi_{ii,1} =$		
	-.3	-.5	-.9
$\bar{\hat{\lambda}}_2(T = 100)$	.023	.011	.001
$\bar{\hat{\lambda}}_2(T = 500)$	.004	.003	0
$\bar{\hat{\lambda}}_2(T = 1000)$	.002	.001	0
$\bar{\hat{\lambda}}_2(T = 2000)$	.001	0	0
λ_2	0	0	0

$\pi_{31,1} = .5$			
	$\pi_{ii,1} =$		
	-.3	-.5	-.9
$\bar{\hat{\lambda}}_2(T = 100)$	.122	.061	.002
$\bar{\hat{\lambda}}_2(T = 500)$	.108	.052	.001
$\bar{\hat{\lambda}}_2(T = 1000)$	.106	.051	0
$\bar{\hat{\lambda}}_2(T = 2000)$	.105	.05	0
λ_2	.104	.049	.006

the selection rule underestimates the rank of $\Sigma_{g(\hat{\pi})}$ for this data generating process.[3] However, despite the incorrect decision of the selection rule, the empirical size of the Wald test with generalized inverse falls inside the 2–standard–deviation bound around 5%, at least in large samples (see Figures 6.6 (c), (e) and (g)).

In Figure 6.5 (b), $\pi_{31,1} = 0$ and the true rank of $\Sigma_{g(\hat{\pi})}$ equals one. This rank is recognized for any ϵ value and any sample size. Therefore, the empirical size hardly varies for different ϵ values in Figures 6.6 (b), (d), (f) and (h). Moreover, all figures show that the Wald test with generalized inverse tends to have a large type I error for the present DGP.

Results can be summarized as follows:

For the data generating process of Model I, the randomized Wald test has good empirical size close to the nominal size of 5% if $\xi \in [.1,\ .5]$.

The selection rule correctly estimates the rank of $\Sigma_{g(\hat{\pi})}$ for processes well inside the stationary border ($\pi_{ii,1} \in \{-.3, -.5\}$), if $\epsilon \in [.5, .7]$ where the larger value should be preferred in small samples.

However, even if the selection rule underestimates the rank of $\Sigma_{g(\hat{\pi})}$, the Wald test with generalized inverse has empirical size well inside the 2–standard–deviation bound around 5%.

Simulation of Power for Model I

To simulate the power, data generating processes have been considered which do not fulfill the null hypothesis. In particular, the coefficient $\pi_{21,1}$ has been chosen equal to $\delta_1/\sqrt{T}$ in Model I. Since $\pi_{23,1} = 0$,

$$\begin{aligned} \pi^{(2)}_{21,1} &= (\pi_{11,1} + \pi_{22,1})\pi_{21,1} + \pi_{23,1}\pi_{31,1} \\ &= (\pi_{11,1} + \pi_{22,1})\pi_{21,1}, \end{aligned}$$

such that the vector of restrictions is

$$g(\pi) = \begin{bmatrix} \pi_{21,1} \\ \pi^{(2)}_{21,1} \end{bmatrix}$$

[3] The results of Table 6.3 indicate a problem of the selection rule: if the true eigenvalue λ_2 is close to zero, the selection rule may not be able to discover that $\lambda_2 > 0$. This raises the question whether taking into account the scaling of the data will improve the performance of the selection rule.

$$
\begin{aligned}
&= \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_1 \\ (\pi_{11,1} + \pi_{22,1})\delta_1 \end{bmatrix} \\
&= \frac{1}{\sqrt{T}} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} \neq 0
\end{aligned}
$$

for $\delta_1 \neq 0$.

The power function $F(\delta_1) = P(\text{reject } H_0^{\text{EGC}}|\delta_1)$ shows the probability to reject the null hypothesis H_0^{EGC} for a given value δ_1. The power function has been computed for $\delta_1 = \{0, .1, .2, \ldots, 3.9, 4\}$ for the standard Wald statistic $\mathcal{W}_{\text{EGC}}$, the Wald statistic with generalized inverse $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ and the randomized Wald statistics $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01), \ldots, \mathcal{W}^{(\xi)}(\xi = 10)$. Simulation results have been computed for all eigenvalues $\pi_{ii,1} \in \{-.9, -.5, -.3, .3, .5, .9\}$ but are reported for negative eigenvalues only. Note, that results obtained with positive eigenvalues draw the same picture.

If $\delta = [\delta_1, \delta_2]' \neq 0$, the standard Wald statistic follows asymptotically a $\chi^2(2, \gamma)$–distribution with noncentrality parameter $\gamma = \delta' \Sigma^{-1}_{g(\hat{\pi})} \delta$ under the null hypothesis H_0^{EGC}. The theoretical power function can be computed once γ is known.

Computation of the noncentrality parameter γ requires computation of the covariance matrix $\Sigma_{g(\hat{\pi})} = (\partial g(\pi)/\partial \pi') \Sigma_{\hat{\pi}} (\partial g(\pi)'/\partial \pi)$. The form of the Jacobian matrix $(\partial g(\pi)/\partial \pi')$ has already been derived in Example 4.1:

$$
\begin{aligned}
\frac{\partial g(\pi)}{\partial \pi'} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \pi_{21,1} & \pi_{11,1} + \pi_{22,1} & \pi_{23,1} & 0 & \pi_{21,1} & 0 & 0 & \pi_{31,1} & 0 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ (1/\sqrt{T})\delta_1 & \pi_{11,1} + \pi_{22,1} & 0 & 0 & (1/\sqrt{T})\delta_1 & 0 & 0 & \pi_{31,1} & 0 \end{bmatrix}.
\end{aligned}
$$

Note, that even if $\pi_{31,1} = 0$, the Jacobian matrix has full row rank as long as $\delta_1 \neq 0$ and hence $\pi_{21,1} \neq 0$.

The covariance matrix of the estimated VAR coefficients can be computed as

$$\Sigma_{\hat{\pi}} = \Gamma_y^{-1} \otimes \Sigma_u,$$

with Γ_y the autocovariance matrix of the vector y at lag 0, and $\Sigma_u = I_3$ by assumption

in (6.1). Thereby, Γ_y can be computed from

$$\text{vec}(\Gamma_y) = (I_k^2 - \Pi_1 \otimes \Pi_1)^{-1}\text{vec}(\Sigma_u)$$

(see Lütkepohl (1991, Chapter 2)).

Hence, for a given parameter value δ_1, $(\partial g(\pi)/\partial \pi')$, $\Sigma_{\hat{\pi}}$ and hence $\Sigma_{g(\hat{\pi})}$ can be computed easily. If furthermore the Jacobian matrix $(\partial g(\pi)/\partial \pi')$ has full row rank, $\Sigma_{g(\hat{\pi})}^{-1}$ and hence the noncentrality parameter γ and the value of the power function can be computed. A problem occurs only for $\delta_1 = 0$ and $\pi_{31,1} = 0$ because the Jacobian matrix does not have full row rank for these parameter values.

If the Jacobian matrix $(\partial g(\pi)/\partial \pi')$ has full row rank, the theoretical power function of the Wald statistic with generalized inverse behaves just as the standard Wald statistic, hence it follows asymptotically a noncentral $\chi^2(2, \gamma)$–distribution with γ computed as described above. However, in the special case where $\delta_1 = 0$ and $\pi_{31,1} = 0$, $\Sigma_{g(\hat{\pi})}$ is not invertible. In this case, the Wald statistic with generalized inverse follows asymptotically a noncentral $\chi^2(1, \gamma^+)$–distribution where $\gamma^+ = \delta' \Sigma^+_{g(\hat{\pi})} \delta$ and $\Sigma^+_{g(\hat{\pi})}$ computed as in Example 5.3.

In practice, the power of the Wald statistic with generalized inverse depends on the decision of the selection rule. If the Jacobian matrix has full row rank but the selection rule assumes incorrectly that $\text{rk}(\Sigma_{g(\hat{\pi})}) = 1$, the power function will follow a $\chi^2(1, \gamma^+)$–distribution. Hence, depending on the decision of the selection rule, the power function of the modified Wald statistics $\mathcal{W}^+_{\text{EGC}}(\epsilon = .33), \ldots, \mathcal{W}^+_{\text{EGC}}(\epsilon = 1)$ will be bounded between a $\chi^2(2, \gamma)$–distribution and a $\chi^2(1, \gamma^+)$–distribution, and thus between the power functions of the standard Wald statistic $\mathcal{W}_{\text{EGC}}$ and the Wald statistic $\mathcal{W}^+_{\text{EGC}}(\text{rk} = 1)$.

The power function of the randomized Wald statistic follows asymptotically a noncentral $\chi^2(2, \gamma^{(\xi)})$–distribution with $\gamma^{(\xi)} = \delta' \left[\Sigma_{g(\hat{\pi})} + \xi \Sigma_w\right]^{-1} \delta$ and hence depends on ξ. For the present trivariate VAR(1) model, let

$$\Sigma_{g(\hat{\pi})} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix},$$

be a (2×2) matrix with σ_{11} the variance of $\hat{\pi}_{21,1}$, σ_{12} the covariance of $\hat{\pi}_{21,1}$ and $\hat{\pi}^{(2)}_{21,1}$

and σ_{22} the variance of $\hat{\pi}^{(2)}_{21,1}$, and let

$$\xi\Sigma_w = \begin{bmatrix} 0 & 0 \\ 0 & \xi\sigma_{11} \end{bmatrix} \tag{6.5}$$

be a (2×2) matrix with the variance of $\hat{\pi}_{21,1}$ as lower left element and zeros elsewhere (cf. Example 5.1). Then

$$\Sigma^{-1}_{g(\hat{\pi})} = \frac{1}{\Delta} \begin{bmatrix} \sigma_{22} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}, \tag{6.6}$$

where $\Delta = \sigma_{11}\sigma_{22} - \sigma^2_{12}$, and

$$\left[\Sigma_{g(\hat{\pi})} + \xi\Sigma_w\right]^{-1} = \frac{1}{\tilde{\Delta}} \begin{bmatrix} \sigma_{22} + \xi\sigma_{11} & -\sigma_{12} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}, \tag{6.7}$$

where

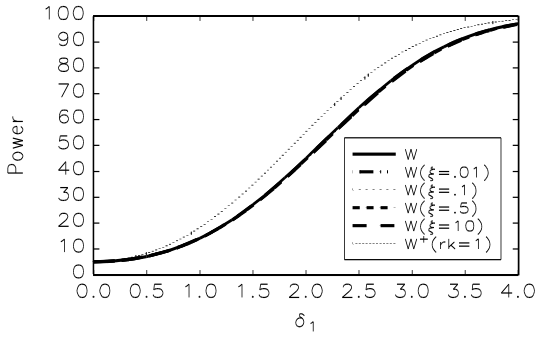
$$\frac{1}{\tilde{\Delta}} = \frac{1}{\Delta}\frac{\Delta}{\Delta + \xi\sigma^2_{11}}. \tag{6.8}$$

It turns out that in the trivariate VAR(1) model, adding noise to the standard Wald statistic affects the power in two different ways: First, the higher the value for ξ, the larger the determinant $\tilde{\Delta}$ relative to Δ. Second, the higher the value for ξ, the more weight is put on δ_1, hence on the violation of the first restriction.

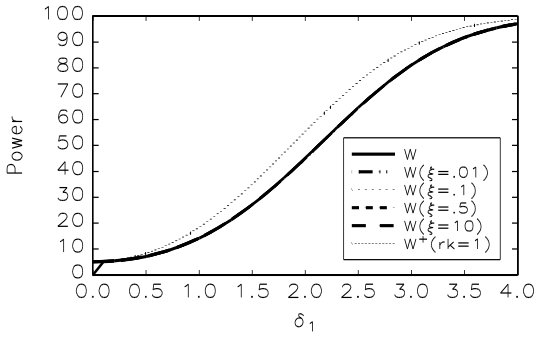
Depending on which effect outweights the other one, the randomized Wald test may have lower power but may also have higher power than the standard Wald test in those cases where the first restriction is violated.

Theoretical power functions for the standard Wald statistic $\mathcal{W}_{\text{EGC}}$, the randomized Wald statistic $\mathcal{W}^{(\xi)}_{\text{EGC}}$ for $\xi \in \{.01, .1, .5, 10\}$ and the Wald statistic with generalized inverse $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ where $\text{rk}(\Sigma_{g(\hat{\pi})}) = 1$ is assumed throughout have been computed for $T = 1000$ and $\delta_1 \in \{0, .1, \ldots, 3.9, 4\}$. These are given in Figure 6.7. Note, that since $\Sigma^{-1}_{g(\hat{\pi})}$ does not exist if $\delta_1 = 0$ and $\pi_{31,1} = 0$, the value of the power function of the standard Wald statistic cannot be computed at this point and has been set to zero:

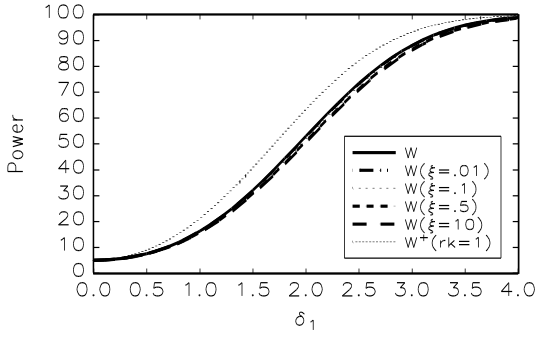
Apart from Figure 6.7 (a), the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ performs best in terms of power. This result does not come as a surprise, given that the Wald statistic $\mathcal{W}^{+}(\text{rk} = 1)$ gives more weight to the first restriction and that the power functions have been computed for a data generating process where the first restriction is violated.



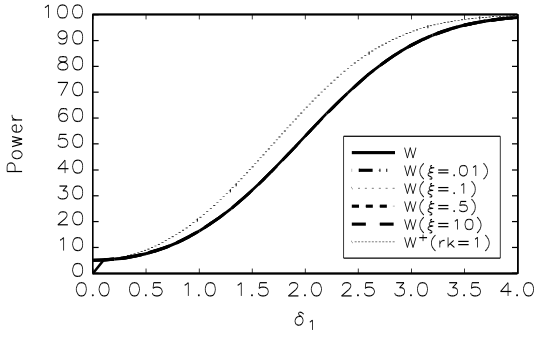
(a) $\pi_{ii,1} = -.3$, $\pi_{31,1} = .5$



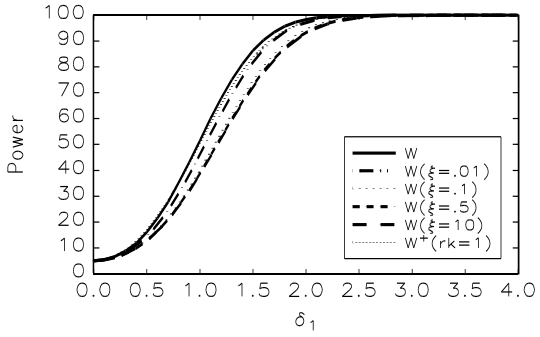
(b) $\pi_{ii,1} = -.3$, $\pi_{31,1} = 0$



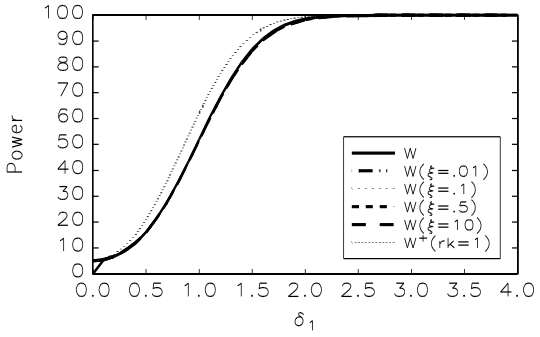
(c) $\pi_{ii,1} = -.5$, $\pi_{31,1} = .5$



(d) $\pi_{ii,1} = -.5$, $\pi_{31,1} = 0$



(e) $\pi_{ii,1} = -.9$, $\pi_{31,1} = .5$



(f) $\pi_{ii,1} = -.9$, $\pi_{31,1} = 0$

Figure 6.7: Theoretical Power Functions of Model I.

Astonishingly, the power functions of the randomized Wald statistics $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$, $\ldots$, $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$ hardly differ for different ξ values and are almost identical to the power function of the standard Wald statistic. These results contradict the expectation that the power decreases with increasing ξ values: only Figure 6.7 (e) points into this direction.

Consider now the simulation results in Tables 6.4 to 6.7: by and large, the power simulations draw the same picture as the theoretical power functions in Figure 6.7. The new information in Tables 6.4 to 6.7 consists of the simulated power of the Wald statistics $\mathcal{W}^{+}(\epsilon)_{\text{EGC}}$ with $\epsilon \in \{.33, .5, .67, .75, .9, 1\}$ for which theoretical power functions could not be computed:

It turns out that if the VAR(1) process is well inside the stationary border ($\pi_{ii,1} \in \{-.3, -.5\}$) and if $\pi_{31,1} = .5$, the selection rule correctly estimates the true rank of $\Sigma_{g(\hat{\pi})}$ for any $\epsilon > .33$. For these ϵ values, the Wald statistics $\mathcal{W}^{+}(\epsilon)_{\text{EGC}}$ have the same power as the standard Wald statistic. They have less power than the modified Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$, but since the latter statistic is based on an incorrect assumption on the true rank of $\Sigma_{g(\hat{\pi})}$, its higher power should not be taken as yardstick here.

If $\pi_{31,1} = 0$, and $\delta \in \{.3, .5, 1\}$, the power of the standard Wald statistic is low. However, for $\delta \in \{.3, .5, 1\}$ and $T = 1000$ observations, $\pi_{21,1} = \delta/\sqrt{T}$ is still close to the singularity point $\pi_{21,1} = 0$. The low power indicates that the standard Wald statistic may not only have problems if the covariance matrix $\Sigma_{g(\hat{\pi})}$ is singular but also if $\Sigma_{g(\hat{\pi})}$ is nearly singular (see also Bolfarine et al. (2001)).

The Wald statistic $\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$ performs much better in terms of power and so do the modified Wald statistics $\mathcal{W}^{+}_{\text{EGC}}(\epsilon)$ for values $\epsilon < .75$ which cause the selection rule to underestimate the rank of $\Sigma_{g(\hat{\pi})}$. Put differently, a correct decision of the selection rule is obtained if ϵ is chosen equal to or larger than .75.

Table 6.6 shows that if $\pi_{ii,1} = -.9$, the rank of $\Sigma_{g(\hat{\pi})}$ is underestimated for any $\epsilon \leq .9$. This confirms the result of Figure 6.5 that the selection rule does not work well if the VAR(1) process is close to the nonstationary border.

In small samples of size $T = 100$, the Wald test with generalized inverse shows good power properties if $\epsilon \in \{.5,\ .75\}$, see Table 6.7.

Table 6.4

Power of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 1000$.

	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} = -.3$											
	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$						$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$					
	0	.3	.5	1	2	4	0	.3	.5	1	2	4
$\mathcal{W}_{\text{EGC}}$	5.0	5.6	6.9	14.0	44.7	97.1	1.6	2.2	3.1	8.6	37.1	96.0
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	5.0	5.7	7.1	14.1	44.8	97.1	3.6	4.3	5.5	11.8	41.5	96.7
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	5.2	6.1	7.3	14.3	44.7	97.0	5.1	6.0	7.3	14.3	44.8	97.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	5.4	6.1	7.4	14.6	44.8	96.9	5.3	6.3	7.5	14.7	45.3	97.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.2	6.2	7.4	14.6	44.7	96.9	5.3	6.3	7.5	14.8	45.4	97.1
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	5.0	5.6	7.1	14.8	46.7	97.4	5.2	6.3	8.5	18.9	55.8	98.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	5.0	5.6	6.9	14.0	44.7	97.1	5.2	6.3	8.5	18.9	55.8	98.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	5.0	5.6	6.9	14.0	44.7	97.1	5.2	6.3	8.4	18.6	55.2	98.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	5.0	5.6	6.9	14.0	44.7	97.1	2.4	3.0	4.2	10.7	40.3	96.3
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	5.0	5.6	6.9	14.0	44.7	97.1	1.6	2.2	3.1	8.6	37.1	96.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	5.0	5.6	6.9	14.0	44.7	97.1	1.6	2.2	3.1	8.6	37.1	96.0
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	5.2	6.3	8.4	18.9	55.5	98.6	5.2	6.3	8.5	18.9	55.8	98.6

Table 6.5

Power of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 1000$.

	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} = -.5$											
	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$						$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$					
	0	.3	.5	1	2	4	0	.3	.5	1	2	4
$\mathcal{W}_{\text{EGC}}$	4.5	5.4	7.0	15.5	52.7	98.9	1.6	2.2	3.2	10.1	45.7	98.4
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	4.6	5.5	7.1	15.6	52.8	99.0	3.9	4.6	5.9	13.9	50.6	98.7
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	4.9	6.0	7.5	15.5	51.9	98.8	5.2	6.0	7.4	16.1	53.6	98.9
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	5.1	6.0	7.6	15.5	51.2	98.7	5.4	6.2	7.5	16.4	53.8	98.9
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.3	6.1	7.3	15.9	50.9	98.6	5.4	6.3	7.6	16.4	53.9	98.9
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	5.1	6.3	8.9	21.8	63.9	99.6	5.0	6.3	8.9	21.9	63.8	99.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	4.5	6.3	7.0	15.5	52.7	98.9	5.0	6.3	8.9	21.9	63.8	99.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	4.5	5.4	7.0	15.5	52.7	98.9	5.0	6.3	8.9	21.9	63.8	99.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	4.5	5.4	7.0	15.5	52.7	98.9	5.0	6.3	8.9	21.9	63.7	99.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	4.5	5.4	7.0	15.5	52.7	98.9	1.6	2.2	3.2	10.2	45.7	98.4
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	4.5	5.4	7.0	15.5	52.7	98.9	1.6	2.2	3.2	10.1	45.7	98.4
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	5.1	6.3	8.9	21.8	63.9	99.6	5.0	6.3	8.9	21.9	63.8	99.6

Table 6.6

Power of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 1000$.

	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} = -.9$											
	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$						$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$					
	0	.1	.3	.5	1	2	0	.1	.3	.5	1	2
$\mathcal{W}_{\text{EGC}}$	5.2	5.3	8.5	15.9	51.1	98.5	1.6	1.9	4.2	10.4	44.2	98.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	5.3	5.3	8.0	14.4	45.7	97.0	4.8	5.1	8.2	15.9	51.5	98.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	4.9	5.1	7.2	12.3	38.9	94.4	5.4	5.6	8.8	16.7	52.7	98.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	5.1	5.2	7.3	12.0	37.7	93.7	5.7	5.7	8.9	16.7	52.9	98.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.1	5.2	7.3	11.9	37.6	93.4	5.4	5.7	8.8	16.7	52.9	98.6
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	4.2	4.4	7.2	13.9	47.5	97.5	5.2	5.7	11.1	21.9	63.5	99.5
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	4.7	5.0	8.7	16.1	49.6	97.5	5.2	5.7	11.1	21.9	63.5	99.5

Table 6.7

Power of Modified Wald Tests for Model I,

Nominal Significance Level 5%, $T = 100$.

	$\pi_{11,1} = \pi_{22,1} = \pi_{33,1} = -.5$											
	$\pi_{31,1} = .5$						$\pi_{31,1} = 0$					
	$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$						$\pi_{21,1} = \delta_1/\sqrt{T}, \quad \delta_1 =$					
	0	.3	.5	1	2	4	0	.3	.5	1	2	4
$\mathcal{W}_{\text{EGC}}$	4.4	4.9	6.4	14.9	50.6	97.8	2.0	2.7	4.1	11.3	45.4	97.3
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	4.8	5.2	6.7	15.2	50.7	97.8	2.3	2.9	4.2	11.9	46.2	97.4
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	5.1	5.7	7.2	15.6	50.5	97.7	4.2	4.8	6.3	14.7	50.0	97.8
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	5.4	6.3	7.7	15.7	50.1	97.6	5.2	5.9	7.5	16.5	52.1	98.0
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.6	6.5	7.7	15.9	50.0	97.6	5.6	6.3	8.1	17.2	52.8	98.1
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	5.8	7.1	9.8	21.8	63.1	99.0	5.8	7.2	9.9	22.0	63.7	99.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	5.3	6.3	8.6	21.5	62.6	98.4	5.7	7.2	9.9	22.0	63.7	99.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	4.5	4.9	6.5	15.9	53.2	97.8	5.3	6.5	9.2	20.4	60.4	98.3
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	4.4	4.9	6.4	14.9	50.8	97.8	3.9	5.1	7.3	17.1	53.9	97.8
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	4.4	4.9	6.4	14.9	50.8	97.8	2.1	2.8	4.2	11.7	46.0	97.3
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	4.4	4.9	6.4	14.9	50.8	97.8	2.1	2.7	4.1	11.3	45.5	97.3
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	5.8	7.1	9.8	21.8	63.1	99.0	5.8	7.2	9.9	22.0	63.7	99.0

In Figure 6.7, the theoretical power functions of the randomized Wald statistics $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$,..., $\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$ do not vary much for different ξ values. In particular, the randomized Wald statistics cannot outperform the standard Wald statistic in terms of power. In contrast, Tables 6.4 and 6.5 show that if $\pi_{31,1} = 0$, the simulated power of the randomized Wald test is significantly higher than the power of the standard Wald test. Moreover, the simulated power of the randomized Wald test slightly increases with increasing ξ value. The latter result may hinge on the specific data generating process where the first restriction is violated. This effect should not arise if only the second restriction in $g(\pi)$ is violated. The power analysis of Model II is based on such a data generating process.

Simulation of Size and Power for Model II

The power analysis of Model I may draw a picture which is too optimistic. Therefore, the power has also been studied for the data generating process of Model II. Setting $\pi_{23,1} = \delta_2/\sqrt{T}$ with $\delta_2 \neq 0$ in (6.4) on page 118 yields a vector

$$g(\pi) = \frac{1}{\sqrt{T}} \begin{bmatrix} 0 \\ 0.7\delta_2 \end{bmatrix} \tag{6.9}$$

where only the second restriction is violated. If $\delta_2 = 0$, $g(\pi) = 0$ fulfills the null hypothesis H_0^{EGC}. Consider first the theoretical power functions for Model II, computed for $\delta_2 = \{0, .1, .2, \ldots, 3.9, 4\}$ and depicted in Figure 6.8.

This figure shows, that the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}$(rk=1) has very unfavorable power properties under the data generating process of Model II. The reason is, that the latter statistic uses a generalized inverse which gives more weight to the first restriction than to the second in the computation of the modified Wald statistic. However, since the first restriction is fulfilled, this strategy leads to a small value of the Wald statistic $\mathcal{W}^{+}_{\text{EGC}}$(rk=1). As consequence, a Wald test which is based on the latter statistic, has no power at all.

The randomized Wald statistic suffers from low power, too, if too much random noise is added to the vector of restrictions. Under the alternative (6.9), the randomized Wald

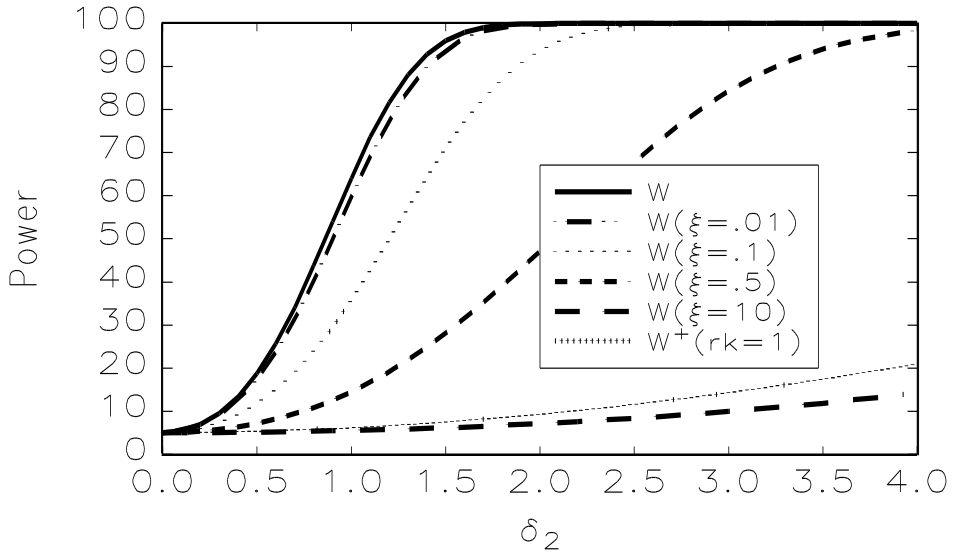


Figure 6.8: Theoretical Power Function of Model II.

statistic $\mathcal{W}_{\text{EGC}}^{(\xi)}$ follows a $\chi^2(2, \gamma^{(\xi)})$–distribution with noncentrality parameter

$$\begin{aligned}\gamma^{(\xi)} &= \delta' \left[\Sigma_{g(\hat{\pi})} + \xi\Sigma_w\right]^{-1} \delta \\ &= (1/\tilde{\Delta})(0.7^2\delta_2^2\sigma_{11}) \\ &= (\Delta/\tilde{\Delta})\gamma,\end{aligned}$$

with γ the noncentrality parameter of the standard Wald statistic, see page 136. Due to the larger determinant $\tilde{\Delta}$, the randomized Wald test will have lower power than the standard Wald test, the larger ξ. Indeed, for $\xi \to \infty$, we have $\gamma^{(\xi)} \to 0$. In other words, the power of the randomized Wald test shrinks towards the size. This effect can be observed in Figure 6.8 already for a ξ value as large as $\xi = 10$.

Figure 6.8 has shown that the strength of the Wald statistic $\mathcal{W}_{\text{EGC}}^{+}$(rk=1), the high power if the first restriction is violated, turns into a weakness for data generating processes where only the second restriction is violated. The performance of the Wald statistics $\mathcal{W}_{\text{EGC}}^{+}(\epsilon)$ for $\epsilon \in \{.33, \ldots, 1\}$ will depend strongly on the correct rank estimation of $\Sigma_{g(\hat{\pi})}$ and hence on the appropriate ϵ value. The simulation results in Tables 6.8 and 6.9 support the conclusions drawn from the theoretical power functions:
First, the randomized Wald test has highest power if $\xi = .01$ but is still less powerful than the standard Wald test. Second, the Wald test with generalized inverse performs poorly if the true rank of $\Sigma_{g(\hat{\pi})}$ is underestimated.

Table 6.8

Size and Power of Modified Wald Tests for Model II,

Nominal Significance Level 5%, $T = 1000$.

	$\pi_{23,1} = \delta_2/\sqrt{T}, \quad \delta_2 =$						
	-.5	0	.3	.5	1	2	4
$\mathcal{W}_{\text{EGC}}$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	18.3	5.3	8.0	15.2	54.0	99.4	100
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	12.8	5.8	7.0	10.5	31.7	90.9	100
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	8.3	5.7	6.1	7.0	13.2	44.1	97.6
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.7	5.4	5.4	5.4	5.8	7.3	13.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	5.9	5.4	5.5	5.5	6.2	8.6	19.3
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	19.0	5.1	8.4	16.5	58.0	99.6	100
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	5.7	5.4	5.5	5.6	6.2	8.6	19.1

However, there is some good news: in large samples ($T = 1000$), Table 6.8 shows that the selection rule correctly estimates the rank of $\Sigma_{g(\hat{\pi})}$ for any value $\epsilon > .33$. In small samples ($T = 100$), good results are obtained for $\epsilon \geq .67$. For these ϵ values, the Wald test with generalized inverse has (almost) the same power as the standard Wald test. Tables 6.8 and 6.9 show that the modified Wald tests also have empirical size well inside the 2–standard–deviation bound around 5% for these ξ and ϵ values.[4]

[4]In Table 6.9, the power has not been simulated for $\delta \geq 2$ because the data generating process of Model II is no longer stationary for these δ_2 values.

Table 6.9

Size and Power of Modified Wald Tests for Model II,

Nominal Significance Level 5%, $T = 100$.

	$\pi_{23,1} = \delta_2/\sqrt{T}$, $\delta_2 =$				
	-.5	0	.3	.5	1
$\mathcal{W}_{\text{EGC}}$	17.3	5.3	5.7	10.0	42.1
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .01)$	16.7	5.2	5.6	9.9	38.7
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .1)$	12.9	5.5	5.7	7.7	23.5
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = .5)$	8.2	5.7	5.6	6.3	11.3
$\mathcal{W}^{(\xi)}_{\text{EGC}}(\xi = 10)$	5.5	5.5	5.4	5.5	5.9
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .33)$	7.1	5.6	5.1	5.0	5.1
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .5)$	14.4	5.2	4.4	5.2	12.1
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .67)$	17.3	5.2	5.2	8.5	32.5
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .75)$	17.3	5.2	5.6	9.7	39.0
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = .9)$	17.3	5.3	5.7	10.0	42.1
$\mathcal{W}^{+}_{\text{EGC}}(\epsilon = 1)$	17.3	5.3	5.7	10.0	42.1
$\mathcal{W}^{+}_{\text{EGC}}(\text{rk} = 1)$	6.8	5.5	5.1	5.0	5.0

Figure 6.10 shows that the Wald test with generalized inverse has empirical size close to the nominal size of 5% even for a sample size $T = 100$. Moreover, Figure 6.9 illustrates once again that in small samples a larger value $\epsilon > .67$ should be preferred to guarantee that the rank of $\Sigma_{g(\hat{\pi})}$ is estimated correctly. For smaller ϵ values, the selection rule tends to underestimate the rank of $\Sigma_{g(\hat{\pi})}$.

All in all, for the DGP of Model II, simulation results speak in favour of a ξ value as small as $\xi = .01$ for the randomized Wald test. For the Wald test with generalized inverse, a value $\epsilon \in [.5, .7]$ should be used for the selection rule with the larger ϵ value

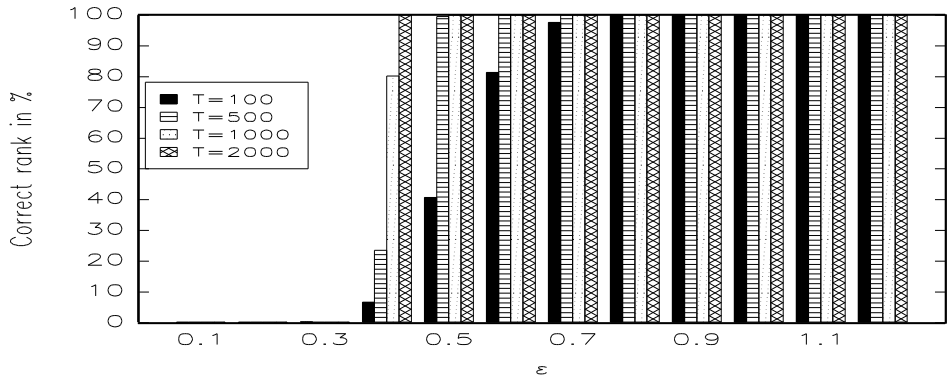
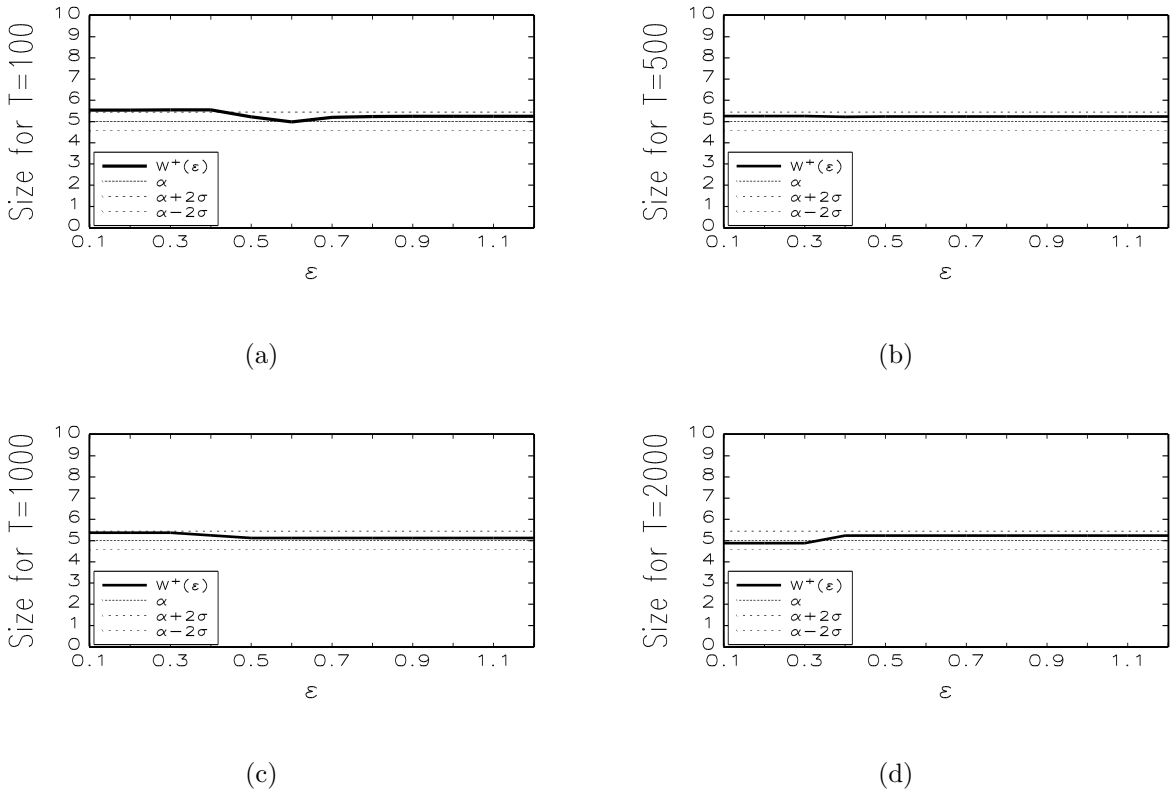


Figure 6.9: Performance of Selection Rule for Model II.



(a) (b)

(c) (d)

Figure 6.10: Size of $\mathcal{W}^{+}_{\mathrm{EGC}}(\epsilon)$ in Model II.

preferred in small samples.

6.3 Conclusions

The simulation study has shown that the standard Wald test suffers from size distortions if the Jacobian matrix $\partial g(\pi)/\partial \pi'$ is evaluated at parameter sets under the null hypothesis which imply a reduced row rank of the Jacobian matrix. Indeed, for the data generating processes used in this simulation study, the standard Wald test behaves conservatively at these problematic parameter points.[5] Moreover, the simulation results show that the standard Wald test also has low power under alternatives which are close to singular parameter points (see also Bolfarine et al. (2001) for similar results).

The manifold examples of Chapter 4 have illustrated that it is difficult, if not impossible, to locate all parameter points under H_0^{EGC} which imply a reduced row rank of the Jacobian matrix and hence size and power distortions of the standard Wald test. It would therefore be favorable to have a modified Wald test which does not suffer from these problems. The simulation results show that the randomized Wald test and the Wald test with generalized inverse can outperform the standard Wald test both in size and power. However, the correct choice of ξ and ϵ is crucial:

A ξ value which is too small may not add enough noise to overcome the singularity of the covariance matrix of the restrictions $\Sigma_{g(\hat{\pi})}$ in the problematic case where the Jacobian matrix has reduced row rank. In contrast, a ξ value which is too large leads to a Wald statistic which is dominated by the noise effect, and therefore has no power. Based on the results of the simulation study, a value $\xi = .1$ seems to be a good compromise.

For ϵ too small, the selection rule tends to underestimate the rank of $\Sigma_{g(\hat{\pi})}$, implying low power for some data generating processes (see Model II). In contrast, if ϵ is chosen too large, the selection rule tends to overestimate the rank of $\Sigma_{g(\hat{\pi})}$. In this case, the

[5] Gaffke et al. (1999) show for the special case of a single restriction under the null hypothesis that the limiting distribution of the standard Wald statistic is bounded from above by a $\chi^2(1)$–distribution if the Jacobian matrix vanishes and hence behaves conservatively. However, to my knowledge, there is no general proof that the standard Wald test always behaves conservatively if the Jacobian matrix has reduced row rank under H_0.

Wald test with generalized inverse behaves just as the standard Wald test. Hence, it does not overcome the size and power distortions of the latter in the problematic case where the Jacobian matrix has reduced row rank. To avoid these effects, $\epsilon \in [.5, .7]$ seems to be a good choice for the data generation processes considered here. In small samples, the larger ϵ value should be preferred.
Note, that choosing a value $\epsilon = .7$ in small samples does not contradict assumption (5.12) since the latter assumption describes the limiting behaviour of a sequence of random variables which need not hold in small samples.

In Chapter 5, a modified version of the Wald test with generalized inverse has been discussed (see Figure 5.1). The simulation results show that a drawback of this strategy is the sensitivity of the selection rule with respect to the size of ϵ. For instance, Figure 6.3 (a) reveals that for $\epsilon = .7$ and $T = 100$, there is a 28% probability that the true rank of $\Sigma_{g(\hat{\pi})}$ is underestimated. Relying on the selection rule only may therefore lead to a large Type I error. However, if the Wald test with generalized inverse is used in the second step, this weakens the Type I error because the latter test has correct asymptotic size as long as the true rank of $\Sigma_{g(\hat{\pi})}$ is not overestimated.

The simulation results have been obtained for the most simple case of a trivariate, stationary VAR model of known lag order $p = 1$. Varying the number of variables or the lag length will increase the number of restrictions which characterize Granger noncausality at all forecast horizons: The more restrictions are tested, the more noise has to be added to the randomized Wald statistic. This may negatively affect the power of the latter statistic.
On the other hand, the higher the number of restrictions, the more likely that the rank of $\Sigma_{g(\hat{\pi})}$ may be under– or overestimated, implying the possibility of size and power distortions of the Wald statistic with generalized inverse.
Moreover, for nonstationary VAR models, the modified Wald tests have to be based on an overfitted VAR model. Since estimating a redundant lag increases the estimation inefficiency, the modified Wald tests may have lower power in nonstationary VAR models than in stationary VAR model.

For these reasons, the present simulation study may draw a picture too optimistic for higher–dimensional, higher–order and/or nonstationary VAR models. However, often causal analyses are limited to bivariate or trivariate models.[6] If the data generating process can be well described by a stationary, trivariate VAR(1) model, the present simulation study shows that the randomized Wald test or the Wald test with generalized inverse should be preferred over the standard Wald test when testing for Granger noncausality at all forecast horizons $h \geq 1$ or for zero impulse responses.

[6]Standard Granger causality has often been studied in bivariate VAR systems, see for instance the large bulk of papers on the relationship between money and outcome. Causality between three variables has been studied e.g. in Hsiao (1979), Drobny & Gausden (1988), Bruneau & Nicolai (1992b), Alexander (1993), Toda & Phillips (1994), Caporale & Pittis (1996), Caporale et al. (1998), Renault et al. (1998) and Giles (2000).

Chapter 7

Summary and Conclusions

Nowadays, Granger causality tests and impulse response analysis are standard tools to investigate causal relationships between variables in a vector autoregressive (VAR) framework. The extensive research in this area has shown that the causal relationship between two variables is not invariant to the information in the VAR system. In particular, it may change if other variables enter the system. This thesis addresses the question whether variable y_1 is causal for variable y_2, given that there is a vector of third variables y_3? This question is answered within the framework of stationary and nonstationary linear vector autoregressive discrete time processes. The notation and different representations of the VAR models used here are given in Chapter 2.

In Chapter 3, the concepts of Granger causality and impulse response analysis are defined within the framework set up in Chapter 2. Moreover, an extended concept of Granger causality is introduced which in contrast to the standard Granger causality concept takes into account indirect causal chains. The concept gains relevance if causality is studied in a multivariate VAR model with three or more variables and may be regarded as a generalization of Granger causality and impulse response analysis (see Dufour & Renault (1998)). To ease the understanding of the notation, several examples have been used. They also help to illustrate the difference between all three causality concepts, which is the key interest of this chapter.

Chapter 4 is directed towards estimation and testing: commonly Wald tests are pre-

ferred to Lagrange Multiplier or Likelihood Ratio tests in testing for causality as they do not require estimation under the null hypothesis. This is especially convenient if nonlinear restrictions are imposed under the null. Nonlinear restrictions on the vectorautoregressive coefficients arise for instance with impulse response analysis. They also arise with a test of Granger noncausality at higher forecast horizons $h > 1$.

Although Wald tests are easy to set up, they may not have their usual asymptotic χ^2–distribution if nonlinear restrictions of the form $g(\pi) = 0$ are tested with π a vector of coefficients and $g(\pi)$ a nonlinear function of the coefficients in π. Problems occur for instance, if the matrix of first order partial derivatives $\partial g(\pi)/\partial \pi'$ does not have full row rank. In this case, the standard Wald statistic may fail to have its usual limiting χ^2–distribution. This problem is commonly neglected in practice. However, Example 4.1, Proposition 4.1 and Corollary 4.1 illustrate that this problem is not irrelevant, in particular for tests of extended Granger causality and impulse response analysis. Moreover, it is illustrated that this problem also affects Likelihood Ratio or Lagrange Multiplier tests.

A similar problem occurs in nonstationary levels VAR models: the coefficient covariance matrix $\Sigma_{\hat{\pi}}$ is singular so that even Wald tests of linear restrictions may now have a nonstandard asymptotic distribution. This problem has already received much attention in the context of standard Granger causality tests. Chapter 4 reviews the literature and discusses whether solutions carry over to tests of extended Granger causality and impulse response analysis. It turns out that solely the idea of Toda & Yamamoto (1995) and Dolado & Lütkepohl (1996) to overfit the true VAR lag length p by d extra lags, but to test restrictions only on the first p VAR coefficient matrices, works for tests of extended Granger causality and impulse response analysis. However, their solutions cannot solve the problem of possibly nonstandard distribution which still arises if the Jacobian matrix $\partial g(\pi)/\partial \pi'$, evaluated at the true parameter values, has a reduced row rank.

Solutions for the latter problem are given in Chapter 5: Propositions 5.1 and 5.2 describe two modified Wald tests, a randomized Wald test and a Wald test with generalized inverse, which both have correct asymptotic size, also in those situations where

the standard Wald statistic may suffer from size distortions:
While the randomized Wald test solves the problem of possibly singular matrix $\Sigma_{g(\hat{\pi})}$ by rendering the estimation less efficient, the Wald test with generalized inverse takes the opposite direction and tries to use the information more efficiently than the standard Wald test. In particular, the information is used that some of the eigenvalues of $\Sigma_{g(\hat{\pi})}$ may be zero under the null hypothesis. A selection rule is used to provide this information: the selection rule sets to zero the eigenvalues of $\Sigma_{g(\hat{\pi})}$ if they are smaller than a prespecified threshold value.
The solutions presented in Propositions 5.1 and 5.2 apply to a test of extended Granger causality in a stationary VAR model. However, Corollary 5.1 and Corollary 5.2 show how results carry over to impulse response analysis under slight modifications. Indeed, the solutions given in this chapter may hold for a variety of other null hypotheses of nonlinear restrictions with similar structure. The extension to nonstationary VAR models is discussed at the end of Chapter 5.
The main attention of Chapter 5 focuses on modifying the standard Wald statistic in a way that the modified Wald statistic has a known limiting distribution under the null hypothesis $H_0 : g(\pi) = 0$ even in the case where the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has reduced row rank. The solutions presented in Propositions 5.1 and 5.2 have therefore been developed in the spirit of Wald tests. This has the advantage that the modified Wald tests have correct asymptotic size. Hence, at least in large samples, the empirical size of the test should be close to the nominal size. However, if the size of the test is no argument, there exist many other alternative testing strategies, mostly sequential procedures, to test the null hypothesis of no Granger causality at all forecast horizons. For completeness, a short description and discussion of these alternative testing strategies is given in Section 5.1.3.

In Chapter 6, the size and power of the modified Wald tests relative to the standard Wald test are investigated for a stationary, trivariate VAR(1) model. In the first part of Chapter 6, theoretical considerations on the size and power of the modified Wald tests as well as theoretical power functions are given for a general trivariate VAR(1) model. The simulation setup and simulation results are then presented in the second

part for two different trivariate VAR(1) models.

The simulation of size is set up in a way which allows to analyze the performance of the standard and the modified Wald statistics for processes where the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has full row rank and others where the row rank of the Jacobian matrix is reduced. The simulation results show that the modified Wald tests perform as well as the standard Wald test, which has empirical size close to the nominal size of 5% if the Jacobian matrix has full row rank. In contrast, if the Jacobian matrix has reduced row rank, the standard Wald test behaves conservatively whereas the modified Wald tests have empirical size much closer to the nominal size of 5%, at least for $\xi \in \{.1, .5\}$, for $\epsilon \in [.5, .7]$ and for processes well inside the stationary region.

The empirical power has been simulated for two different models which can be understood as best case scenario (Model I) and worst case scenario (Model II). The simulation results for Model I show that the Wald test with generalized inverse can outperform the standard Wald test in terms of power. Moreover, although the randomized Wald test is inefficient relative to the standard Wald test, this inefficiency does not show up in a loss in power in Model I. However, the simulation results for Model II correct this optimistic picture: indeed, both modified Wald tests can show a substantial loss in power relative to the standard Wald statistic if too much noise is added, or if the selection rule is based on a threshold value which has been chosen too large (ϵ has been chosen too small).

In practice, the user will not know whether the true DGP is rather of type "Model I" or of type "Model II". Based on the simulation results, a value $\xi = .1$ and a value $\epsilon \in [.5, .7]$ are recommended to ensure an empirical size close to the nominal size even in the problematic case where the Jacobian matrix has reduced row rank as well as to keep the loss in power small relative to the power of the standard Wald statistic.

The present thesis shows two ways of solving the problem that the standard Wald statistic may have an unknown limiting distribution for some parameter values under the null hypothesis if nonlinear restrictions of the kind $g(\pi) = 0$ are involved: a randomized Wald test and a Wald test with generalized inverse. Both modified Wald

tests can outperform the standard Wald test in the stationary trivariate VAR(1) model considered in the simulation study. However, their performance in higher–dimensional, higher–order VAR models, and in particular in nonstationary VAR models, is still open to future research.

Appendix A

Proposition 4.1: An Illustrative Example

An illustration of the decomposition $g(\pi) = G(\pi)\tilde{g}(\pi)$ used in the proof of Proposition 4.1 is given for a VAR(p) model with $k_1 = k_2 = k_3 = 1$ and $p = 3$. For this model, the null hypothesis that y_1 is never Granger causal for y_2 imposes the following set of restrictions:

$$H_0 : \; g(\pi) \;=\; \begin{bmatrix} \pi_{21,1} \\ \pi_{21,2} \\ \pi_{21,3} \\ \pi^{(2)}_{21,1} \\ \pi^{(2)}_{21,2} \\ \pi^{(2)}_{21,3} \\ \pi^{(3)}_{21,1} \\ \pi^{(3)}_{21,2} \\ \pi^{(3)}_{21,3} \\ \pi^{(4)}_{21,1} \\ \pi^{(4)}_{21,2} \\ \pi^{(4)}_{21,3} \end{bmatrix} = 0.$$

Using the recursion formula (3.17), the nonlinear restrictions can be alternatively written as

$$\pi^{(2)}_{21,1} \;=\; \pi_{21,2} + \pi_{21,1}\pi_{11,1} + \pi_{22,1}\pi_{21,1} + \pi_{23,1}\pi_{31,1},$$

$$
\begin{aligned}
\pi^{(2)}_{21,2} &= \pi_{21,3} + \pi_{21,1}\pi_{11,2} + \pi_{22,1}\pi_{21,2} + \pi_{23,1}\pi_{31,2},\\
\pi^{(2)}_{21,3} &= \pi_{21,1}\pi_{11,3} + \pi_{22,1}\pi_{21,3} + \pi_{23,1}\pi_{31,3},
\end{aligned}
$$

$$
\begin{aligned}
\pi^{(3)}_{21,1} &= \pi^{(2)}_{21,2} + \pi^{(2)}_{21,1}\pi_{11,1} + \pi^{(2)}_{22,1}\pi_{21,1} + \pi^{(2)}_{23,1}\pi_{31,1},\\
&= \left((\pi_{11,1} + \pi_{22,1})\pi_{11,1} + \pi_{11,2} + \pi^{(2)}_{22,1}\right)\pi_{21,1} + (\pi_{11,1} + \pi_{22,1})\pi_{21,2}\\
&\quad +\pi_{21,3} + \pi_{11,1}\pi_{23,1}\pi_{31,1} + \pi_{23,1}\pi_{31,2} + \pi^{(2)}_{23,1}\pi_{31,1},\\
\pi^{(3)}_{21,2} &= \pi^{(2)}_{21,3} + \pi^{(2)}_{21,1}\pi_{11,2} + \pi^{(2)}_{22,1}\pi_{21,2} + \pi^{(2)}_{23,1}\pi_{31,2},\\
&= \left((\pi_{11,1} + \pi_{22,1})\pi_{11,2} + \pi_{11,3}\right)\pi_{21,1} + (\pi_{11,2} + \pi^{(2)}_{22,1})\pi_{21,2}\\
&\quad +\pi_{22,1}\pi_{21,3} + \pi_{11,2}\pi_{23,1}\pi_{31,1} + \pi_{23,1}\pi_{31,3} + \pi^{(2)}_{23,1}\pi_{31,2},\\
\pi^{(3)}_{21,3} &= \pi^{(2)}_{21,1}\pi_{11,3} + \pi^{(2)}_{22,1}\pi_{21,3} + \pi^{(2)}_{23,1}\pi_{31,3},\\
&= \left((\pi_{11,1} + \pi_{22,1})\pi_{11,3}\right)\pi_{21,1} + \pi_{11,3}\pi_{21,2}\\
&\quad +\pi^{(2)}_{22,1}\pi_{21,3} + \pi_{11,3}\pi_{23,1}\pi_{31,1} + \pi^{(2)}_{23,1}\pi_{31,3},
\end{aligned}
$$

$$
\begin{aligned}
\pi^{(4)}_{21,1} &= \pi^{(3)}_{21,2} + \pi^{(3)}_{21,1}\pi_{11,1} + \pi^{(3)}_{22,1}\pi_{21,1} + \pi^{(3)}_{23,1}\pi_{31,1},\\
&= \left((\pi_{11,1} + \pi_{22,1})(\pi_{11,1}\pi_{11,1} + \pi_{11,2}) + (\pi_{11,2} + \pi^{(2)}_{22,1})\pi_{11,1} + \pi^{(3)}_{22,1}\right)\pi_{21,1}\\
&\quad +\Big((\pi_{11,1} + \pi_{22,1})\pi_{11,1} + \pi_{11,2} + \pi^{(2)}_{22,1}\Big))\pi_{21,2} + (\pi_{11,1} + \pi_{22,1})\pi_{21,3}\\
&\quad +(\pi_{11,1}\pi_{11,1} + \pi_{11,2})\pi_{23,1}\pi_{31,1} + \pi_{11,1}\pi_{23,1}\pi_{31,2} + \pi_{23,1}\pi_{31,3}\\
&\quad +\pi_{11,1}\pi^{(2)}_{23,1}\pi_{31,1} + \pi^{(2)}_{23,1}\pi_{31,2} + \pi^{(3)}_{23,1}\pi_{31,1},\\
\pi^{(4)}_{21,2} &= \pi^{(3)}_{21,3} + \pi^{(3)}_{21,1}\pi_{11,2} + \pi^{(3)}_{22,1}\pi_{21,2} + \pi^{(3)}_{23,1}\pi_{31,2},\\
&= \left((\pi_{11,1} + \pi_{22,1})(\pi_{11,1}\pi_{11,2} + \pi_{11,3}) + (\pi_{11,2} + \pi^{(2)}_{22,1})\pi_{11,2}\right)\pi_{21,1}\\
&\quad +\left((\pi_{11,1} + \pi_{22,1})\pi_{11,2} + \pi_{11,3} + \pi^{(3)}_{22,1}\right)\pi_{21,2} + (\pi^{(2)}_{22,1} + \pi_{11,2})\pi_{21,3}\\
&\quad +(\pi_{11,1}\pi_{11,2} + \pi_{11,3})\pi_{23,1}\pi_{31,1} + \pi_{11,2}\pi_{23,1}\pi_{31,2}\\
&\quad +\pi_{11,2}\pi^{(2)}_{23,1}\pi_{31,1} + \pi^{(2)}_{23,1}\pi_{31,3} + \pi^{(3)}_{23,1}\pi_{31,2},\\
\pi^{(4)}_{21,3} &= \pi^{(3)}_{21,1}\pi_{11,3} + \pi^{(3)}_{22,1}\pi_{21,3} + \pi^{(3)}_{23,1}\pi_{31,3},\\
&= \left((\pi_{11,1} + \pi_{22,1})\pi_{11,1}\pi_{11,3} + (\pi_{11,2} + \pi^{(2)}_{22,1})\pi_{11,3}\right)\pi_{21,1} + (\pi_{11,1} + \pi_{22,1})\pi_{11,3}\pi_{21,2}\\
&\quad +(\pi_{11,3} + \pi^{(3)}_{22,1})\pi_{21,3} + \pi_{11,1}\pi_{11,3}\pi_{23,1}\pi_{31,1} + \pi_{11,3}\pi_{23,1}\pi_{31,2}\\
&\quad +\pi_{11,3}\pi^{(2)}_{23,1}\pi_{31,1} + \pi^{(3)}_{23,1}\pi_{31,3}.
\end{aligned}
$$

The set of restrictions can be rewritten as

$$g(\pi) = G(\pi)\tilde{g}(\pi)$$

where

$$G(\pi) = \begin{bmatrix} I_3 & 0 & 0 & 0 \\ G_{21}(\pi) & I_3 & 0 & 0 \\ G_{31}(\pi) & G_{32}(\pi) & I_3 & 0 \\ G_{41}(\pi) & G_{42}(\pi) & G_{43}(\pi) & I_3 \end{bmatrix},$$

with

$$G_{21}(\pi) = \begin{bmatrix} \pi_{11,1}+\pi_{22,1} & 1 & 0 \\ \pi_{11,2} & \pi_{22,1} & 1 \\ \pi_{11,3} & 0 & \pi_{22,1} \end{bmatrix},$$

$$G_{31}(\pi) = \begin{bmatrix} (\pi_{11,1}+\pi_{22,1})\pi_{11,1}+\pi_{11,2}+\pi^{(2)}_{22,1} & \pi_{11,1}+\pi_{22,1} & 1 \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,2}+\pi_{11,3} & \pi_{11,2}+\pi^{(2)}_{22,1} & \pi_{22,1} \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,3} & \pi_{11,3} & \pi^{(2)}_{22,1} \end{bmatrix},$$

$$G_{32}(\pi) = \begin{bmatrix} \pi_{11,1} & 1 & 0 \\ \pi_{11,2} & 0 & 1 \\ \pi_{11,3} & 0 & 0 \end{bmatrix},$$

$$\mathrm{vec}\left(G_{41}(\pi)\right) = \begin{bmatrix} (\pi_{11,1}+\pi_{22,1})(\pi_{11,1}\pi_{11,1}+\pi_{11,2})+(\pi_{11,2}+\pi^{(2)}_{22,1})\pi_{11,1}+\pi^{(3)}_{22,1} \\ (\pi_{11,1}+\pi_{22,1})(\pi_{11,1}\pi_{11,2}+\pi_{11,3})+(\pi_{11,2}+\pi^{(2)}_{22,1})\pi_{11,2} \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,1}\pi_{11,3}+(\pi_{11,2}+\pi^{(2)}_{22,1})\pi_{11,3} \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,1}+\pi_{11,2}+\pi^{(2)}_{22,1} \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,2}+\pi_{11,3}+\pi^{(3)}_{22,1} \\ (\pi_{11,1}+\pi_{22,1})\pi_{11,3} \\ \pi_{11,1}+\pi_{22,1} \\ \pi_{11,2}+\pi^{(2)}_{22,1} \\ \pi_{11,3}+\pi^{(3)}_{22,1} \end{bmatrix},$$

$$G_{42}(\pi) = \begin{bmatrix} \pi_{11,1}\pi_{11,1}+\pi_{11,2} & \pi_{11,1} & 1 \\ \pi_{11,1}\pi_{11,2}+\pi_{11,3} & \pi_{11,2} & 0 \\ \pi_{11,1}\pi_{11,3} & \pi_{11,3} & 0 \end{bmatrix},$$

$G_{43}(\pi) = G_{32}$ and

$$\tilde{g}(\pi) = \begin{bmatrix} \pi_{21,1} \\ \pi_{21,2} \\ \pi_{21,3} \\ \pi_{23,1}\pi_{31,1} \\ \pi_{23,1}\pi_{31,2} \\ \pi_{23,1}\pi_{31,3} \\ \pi^{(2)}_{23,1}\pi_{31,1} \\ \pi^{(2)}_{23,1}\pi_{31,2} \\ \pi^{(2)}_{23,1}\pi_{31,3} \\ \pi^{(3)}_{23,1}\pi_{31,1} \\ \pi^{(3)}_{23,1}\pi_{31,2} \\ \pi^{(3)}_{23,1}\pi_{31,3} \end{bmatrix}.$$

Appendix B

Bootstrapping Critical Values for the Wald Statistic

Consider the following stable, stationary trivariate VAR(1) model:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \\ y_{3,t} \end{bmatrix} = \begin{bmatrix} .3 & 0 & 0 \\ 0 & .3 & 0 \\ \pi_{31} & .5 & .3 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix}, \tag{B.1}$$

with $\pi_{31,1} \in \{0,\ .5\}$. Since $\pi_{21,1} = \pi_{23,1} = 0$, it follows from Example 4.1, page 67, that the standard Wald statistic for a test of the null hypothesis that y_1 is never Granger causal for y_2 has a limiting $\chi^2(2)$–distribution under H_0^{EGC} if $\pi_{31,1} = .5$. However, if $\pi_{31,1} = 0$, the Jacobian matrix $\partial g(\pi)/\partial \pi'$ has reduced row rank equal to one and the limiting distribution of the standard Wald statistic may be nonstandard. A small simulation has been set up to answer the question whether bootstrapping the standard Wald statistic can approximate the finite sample distribution of the standard Wald statistic in the latter case?

The simulation is set up as follows:
In Step 1, $T + B$ observations are drawn for a k–variate residual series ($k = 3$) from a standard normal distribution. With the so obtained $(k(T + B) \times 1)$ vector $u = [u'_{-B+1},\ u'_{-B+2}, \ldots, u'_0,\ u'_1, \ldots, u'_T]'$ $(T+B)$ data for the trivariate series y_t are generated according to (B.1) with starting values set to zero. The first $B = 100$ presample values are then cut off to eliminate the starting–up effects.

In Step 2, a VAR(1) model with intercept is fitted to the data by ordinary least squares regression, yielding

$$y_t = \hat{\nu} + \hat{\Pi}_1 y_{t-1} + \hat{u}_t, \quad t = 1, \ldots, T. \tag{B.2}$$

Based on the least squares estimate $\hat{\pi} = \text{vec}(\hat{\Pi}_1)$, the standard Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ for a test of the null hypothesis H_0^{EGC}:"y_1 is never Granger causal for y_2" is computed as in (4.11), page 66.

In Step 3, the $(kT \times 1)$ residual vector $\hat{u} = [\hat{u}_1', \ldots, \hat{u}_T']'$ is copied S times into a long vector $\hat{\mathbf{u}}$ from which S new residual series of length T are generated. Each new residual series $\hat{u}^{(s)},\ s = 1, \ldots, S$, is generated by drawing T times without replacement a $(k \times 1)$ unit $\hat{u}_t$ from $\hat{\mathbf{u}}$. This procedure called "balanced bootstrap" has been proposed by Gleason (1988). It ensures that all T subvectors $\hat{u}_t,\ t = 1, \ldots, T$ of the original residual series $\hat{u}$ appear with equal probability over the entire S bootstrap samples.

For each bootstrap residual series $\hat{u}^{(s)},\ s = 1, \ldots, S$, new data are generated in Step 4 as

$$y_t^\star = \hat{\nu} + \hat{\Pi}_1 y_{t-1}^\star + \hat{u}_t^{(s)}, \quad t = 1, \ldots, T. \tag{B.3}$$

On the basis of these data, a VAR(1) model is fitted by ordinary least squares yielding the estimates $\hat{\hat{\Pi}}_1,\ \hat{\hat{\pi}}_1 = \text{vec}(\hat{\hat{\Pi}}_1)$ and $\hat{\hat{u}}_t^{(s)}$.

The standard Wald statistic is then computed in Step 5 as

$$\mathcal{W}_{\text{EGC}}^{(s)}(\tilde{h}) \;=\; T(g(\hat{\hat{\pi}}) - g(\hat{\pi}))' \hat{\hat{\Sigma}}_{g(\hat{\pi})}^{-1} (g(\hat{\hat{\pi}}) - g(\hat{\pi})), \tag{B.4}$$

where $g(\hat{\hat{\pi}})$ and $\hat{\hat{\Sigma}}_{g(\hat{\pi})}$ are computed on the basis of the estimates of Step 3 while $g(\hat{\pi})$ is the vector of restrictions computed on the basis of the estimates of Step 2. Following a suggestion of Hall & Wilson (1991), the Wald statistic is based on the distance $(g(\hat{\hat{\pi}}) - g(\hat{\pi}))$ instead of the distance of $g(\hat{\hat{\pi}})$ to the true vector $g(\pi) = 0$ under the null hypothesis to improve the power of the Wald test.

Steps 3 to 5 are performed for all $s = 1, \ldots, S$ bootstrap series, yielding a series of S Wald statistics $\mathcal{W}_{\text{EGC}}^{(s)}(\tilde{h})$.

In Step 6, this series is ordered and the value of the Wald statistic $\mathcal{W}_{\text{EGC}}^{(\bar{s})}(\tilde{h})$ with $\bar{s} = (1 - \alpha)S$ is taken as critical bootstrap value. The value of the standard Wald statistic obtained in Step 2 is then compared to this critical value. The null hypothesis

Table B.1

Relative Rejection Frequency $\bar{N}/N$

(Bootstrapped Critical Values)

$\bar{N}/N$		
T	$\pi_{31,1} = 0$	$\pi_{31,1} = .5$
100	.008 (.024)	.034 (.043)
1000	.007 (.01)	.06 (.054)

is rejected if the value of the standard Wald statistic exceeds the critical value. In the simulation, $\alpha = .05$ has been used.

Steps 1 to 6 are repeated for $n = 1, \ldots, N$ with $N = 1000$ the number of simulations. Let $\bar{N}$ be the number of times the null hypothesis is rejected, then the relative rejection frequency of the standard Wald test is computed as $\bar{N}/N$.

Table B.1 presents the relative rejection frequency of the standard Wald test for two different sample sizes $T = 100,\ 1000$ obtained with bootstrapped critical values. The relative rejection frequency of the standard Wald test obtained with the critical value of a $\chi^2(2)$–distribution is given in parentheses.

Using critical values from the bootstrap distribution, the researcher hopes to obtain a test which has finite sample size closer to the nominal size as if he had used the critical values of the asymptotic distribution of the standard Wald statistic (see Jeong & Maddala (1993, p. 581)). Indeed, bootstrapping works well in the regular case where the standard Wald statistic has a limiting $\chi^2(2)$–distribution ($\pi_{31,1} = .5$), at least for $T = 1000$. However, in the case of interest ($\pi_{31,1} = 0$), bootstrapping cannot approximate the finite sample distribution of the standard Wald statistic: the bootstrapped critical values lead to a relative rejection frequency which is even smaller than the one obtained with the critical value of the $\chi^2(2)$–distribution and hence even farther away from the nominal size of 5%.

Figures B.1 and B.2 show the density of the standard Wald statistic $\mathcal{W}_{\text{EGC}}(\tilde{h})$ (solid black line)[1], the density of one bootstrapped Wald statistic $\mathcal{W}^{(s)}_{\text{EGC}}(\tilde{h})$ (dotted black line)[2] and the density of the $\chi^2(2)$–distribution (dotted red line) for different sample sizes $T = 100,\ 1000$. Two lessons can be learned from these figures:

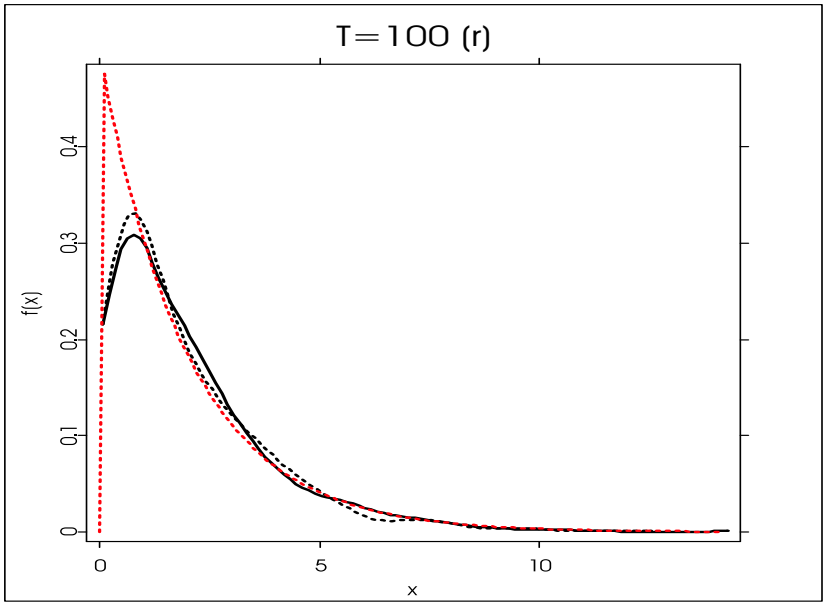
First, Figures B.1 (b) and B.2 (b) show that the standard Wald statistic does not have a limiting $\chi^2(2)$–distribution under H_0^{EGC} if $\pi_{31,1} = 0$ so that $\Sigma_{g(\hat{\pi})}$ is singular. For the present data generation process B.1, the density of the standard Wald statistic has higher probability mass for values in the range between 0 and 2 and less probability mass for values greater than 2 than the $\chi^2(2)$–density. This explains why using a critical value of a $\chi^2(2)$–distribution leads to an empirical size significantly smaller than the nominal size. (See also the simulation results in Tables 6.1 and 6.2).

Second, at first sight the density of the bootstrapped standard Wald statistics describes the density of the standard Wald statistic much better than the $\chi^2(2)$–density. However, a closer look at Figure B.2 (b) reveals that the bootstrap procedure generates bootstrapped Wald statistics with very high values ($x > 15$) which cannot be observed for the original standard Wald statistics series. Due to these outliers, the bootstrapped critical values are too high, leading to the poor results of Table B.1 in the problematic case where $\pi_{31,1} = 0$.

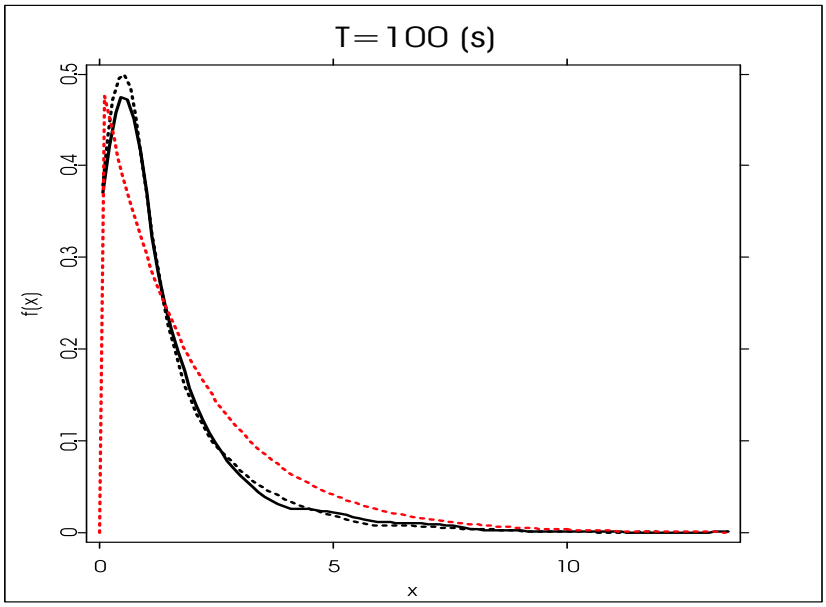
Although Figures B.1 and B.2 show that bootstrapping the density of the standard Wald statistic comes closer to the true density than the density of the $\chi^2(2)$–distribution, at least in the problematic case where $\Sigma_{g(\hat{\pi})}$ is singular, the general performance of the bootstrap procedure is not convincing. For instance, Figure B.2 (b) shows that the bootstrap procedure used here generates outliers which distort the bootstrapped critical values. One reason may be that it was not feasible to bootstrap under the null hypothesis, as recommended in the literature (see e.g. Hall & Wilson (1991)).

[1]The density has been computed from the $N = 1000$ standard Wald statistics using the package `denest` of the statistics software `XploRe`, see Härdle et al. (1999).

[2]The density has been computed from the $S = 1000$ bootstrapped Wald statistics obtained at some step n of the simulation using the package `denest` of the statistics software `XploRe`.

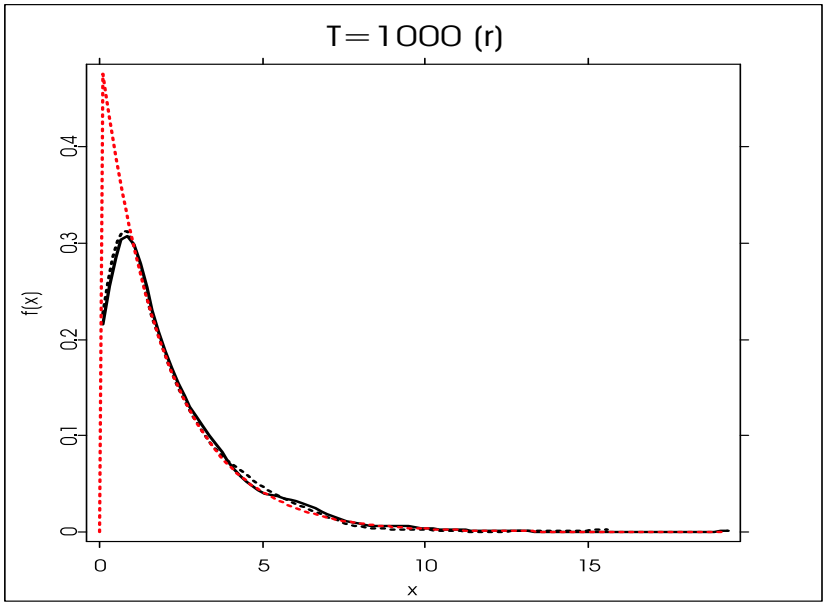


(a) $\pi_{31,1} = .5$

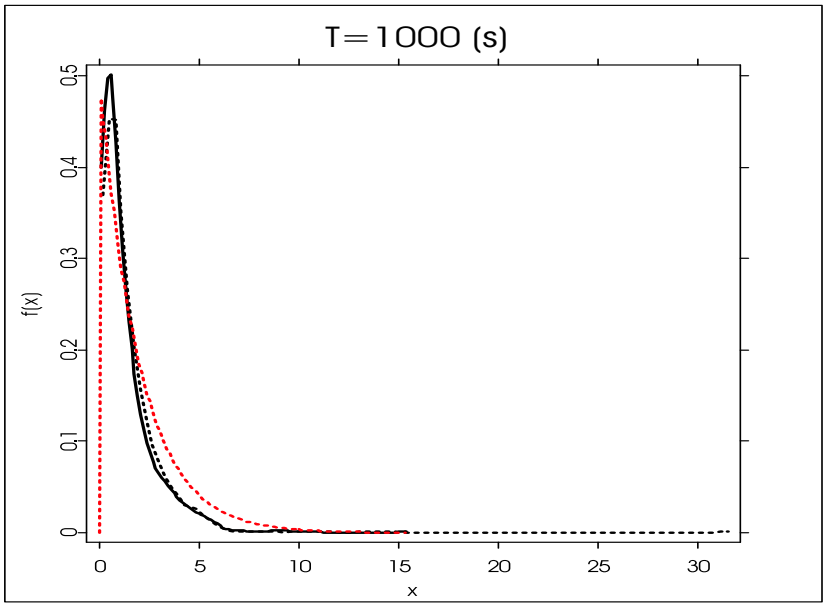


(b) $\pi_{31,1} = 0$

Figure B.1: Density Estimation ($T = 100$).
Density of $\mathcal{W}_{\text{EGC}}(\tilde{h})$: solid black line.
Density of $\mathcal{W}^{(s)}_{\text{EGC}}(\tilde{h})$: dotted black line.
Density of the $\chi^2(2)$–distribution: dotted red line.



(a) $\pi_{31,1} = .5$



(b) $\pi_{31,1} = 0$

Figure B.2: Density Estimation ($T = 1000$).
Density of $\mathcal{W}_{\text{EGC}}(\tilde{h})$: solid black line.
Density of $\mathcal{W}^{(s)}_{\text{EGC}}(\tilde{h})$: dotted black line.
Density of the $\chi^2(2)$–distribution: dotted red line.

Bibliography

Alexander, C. O. (1993). The Changing Relationship Between Productivity, Wages and Unemployment in the UK. *Oxford Bulletin of Economics and Statistics*, 55:87–102.

Amisano, G. & Giannini, C. (1997). *Topics in Structural VAR Econometrics*. Springer Verlag.

Andrews, D. W. K. (1987). Asymptotic Results for Generalized Wald Tests. *Econometric Theory*, 3:348–58.

Banerjee, A., Dolado, J. J., Galbraith, J. W. & Hendry, D. F. (1993). *Co–Integration, Error Correction, and the Econometric Analysis of Non–Stationary Data*. Oxford University Press.

Basmann, R. L. (1988). Causality Tests and Observationally Equivalent Representations of Econometric Models. *Journal of Econometrics*, 39:69–104.

Beeby, M., Hall, S. & Funke, M. (1995). Long–Run Causality and Structural Change: an Application of P–Star Analysis to the UK's Entry into the ERM. LBS Discussion Paper 19, Centre for Economic Forecasting, London, England.

Benkwitz, A., Lütkepohl, H. & Neumann, M. H. (2000). Problems Related to Confidence Intervals for Impulse Responses of Autoregressive Processes. *Econometric Reviews*, 19:69–103.

Bergman, M. & Warne, A. (1993). Money-Income Causality and the Neutrality of Money. Seminar Paper 557, Institute for International Economic Studies, Stockholm University, Sweden.

Bolfarine, H., Cabral, C. R. B. & Paula, G. A. (2001). Distance Tests under Nonregular Conditions: Applications to the Comparative Calibration Model. forthcoming in: Journal of Statistical Computation and Simulation.

Boudjellaba, H., Dufour, J.-M. & Roy, R. (1992a). Simplified Conditions for Non-Causality Between Vectors in Multivariate ARMA Models. CRDE Working Paper 4092, Université de Montréal, Canada.

Boudjellaba, H., Dufour, J.-M. & Roy, R. (1992b). Testing Causality Between Two Vectors in Multivariate ARMA Models. *Journal of the American Statistical Association*, 87:1082–1090.

Braun, P. A. & Mittnik, S. (1993). Misspecification in Vector Autoregressions and their Effects on Impulse Responses and Variance Decompositions. *Journal of Econometrics*, 59:319–41.

Breitung, J. (1998). Neuere Entwicklungen auf dem Gebiet ökonometrischer Strukturmodelle: Strukturelle Vektorautoregressionen. *Ifo Studien: Zeitschrift für empirische Wirtschaftsforschung*, 44:371–392.

Breitung, J. & Heinemann, M. (1998). Short Run Comovements, Persistent Shocks, and the Business Cycle. *Jahrbücher für Nationalökonomie und Statistik*, 217:436–448.

Breitung, J. & Swanson, N. R. (1998). Temporal Aggregation and Causality in Multiple Time Series Models. SFB 373 Discussion Paper 27, Humboldt Universität, Berlin, Germany.

Breusch, T. S. & Schmidt, P. (1988). Alternative Forms of the Wald Test: How Long is a Piece of String? *Communications in Statistics, Theory and Methods*, 17:2789–2795.

Bruneau, C. & Jondeau, E. (1999). Long–Run Causality, with an Application to International Links Between Long–Term Interest Rates. *Oxford Bulletin of Economics and Statistics*, 61:545–568.

Bruneau, C. & Nicolai, J.-P. (1992a). Probabilistic Foundations of a Causal Analysis in a Stationary Vectorial Autoregressive Model. Working paper, ENSAE-CREST and CDC, Paris, France.

Bruneau, C. & Nicolai, J.-P. (1992b). Persistent Causality and Exogeneity in a Multivariate Non–Stationary System. Working paper, ENSAE-CREST and CDC, Paris, France.

Bruneau, C. & Nicolai, J.-P. (1994). Amélioration à la Prévision et Causalité entre Deux Séries d'un Système Multivarié Autorégressif Stationnaire. *Annales d'Economie et de Statistique*, 36:1–22.

Bruneau, C. & Nicolai, J.-P. (1995). Causalité Persistante Entre Series Non-Stationnaires: Application de L'Etude Comparée des Politiques Monetaires des Pays du G5. *Annales d'Economie et de Statistique*, 40:177–206.

Buse, A. (1982). The Likelihood Ratio, Wald, and Lagrange Multiplier Tests: An Expository Note. *The American Statistician*, 36:153–157.

Caporale, G. M., Hassapis, C. & Pittis, N. (1998). Unit Roots and Long–Run Causality: Investigating the Relationship Between Output, Money and Interest Rates. *Economic Modelling*, 15:91–112.

Caporale, G. M. & Pittis, N. (1996). Granger Causality and Weak Exogeneity in Bivariate and Trivariate Vars. LBS Discussion Paper 12, Centre for Economic Forecasting, London, England.

Caporale, G. M. & Pittis, N. (1997). Causality and Forecasting in Incomplete Systems. *Journal of Forecasting*, 16:425–437.

Choi, I. (1993). Asymptotic Normality of the Least-Squares Estimates for Higher Order Autoregressive Integrated Processes with some Applications. *Econometric Theory*, 9:263–282.

Comte, F. & Renault, E. (1996). Non–Causality in Continuous Time Models. *Econometric Theory*, 12:215–256.

Conway, R. K., Swamy, P. A., Yanagida, J. F. & von zur Muehlen, P. (1984). The Impossibility of Causality Testing. *Agricultural Economics Research*, 36:1–19.

Cooley, T. F. & Leroy, S. (1985). Atheoretical Macroeconomics: A Critique. *Journal of Monetary Economics*, 16:283–308.

Cox, D. R. (1992). Causality: Some Statistical Aspects. *Journal of the Royal Statistical Society A*, 155:291–301.

Critchley, F., Marriot, P. & Salmon, M. (1996). On The Differential Geometry of the Wald Test with Nonlinear Restrictions. *Econometrica*, 64:1213–1222.

Davidson, R. & MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press.

Dolado, J., López-Salido, J. D. & Vega, J. L. (2000). Unemployment and Inflation Persistence in Spain: are there Phillips Trade–Offs? *Spanish Economic Review*, 2:267–291.

Dolado, J. & Lütkepohl, H. (1996). Making Wald Tests Work for Cointegrated VAR Systems. *Econometric Reviews*, 15:369–386.

Drobny, A. & Gausden, R. (1988). Granger–Causality, Real Factor Prices and Employment. *European Economic Review*, 32:1261–1283.

Dufour, J.-M. (1989). Nonlinear Hypotheses, Inequality Restrictions, and Non-Nested Hypotheses: Exact Simultaneous Tests in Linear Regressions. *Econometrica*, 57:335–355.

Dufour, J.-M. (1990). Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors. *Econometrica*, 58:475–494.

Dufour, J.-M. & Renault, E. (1994). Short–Run and Long–Run Causality in Time Series. Working paper, GREMAQ and IDEI, Toulouse, France.

Dufour, J.-M. & Renault, E. (1998). Short-Run and Long-Run Causality in Time Series: Theory. *Econometrica*, 66:1099–1125.

Dufour, J.-M. & Tessier, D. (1992). On the Relationship Between Impulse Response Analysis, Innovation Accounting and Granger Causality. *Economics Letters*, 42:327–333.

Eberts, R. W. & Steece, B. (1984). A Test for Granger Causality in a Multivariate ARMA Model. *Empirical Economics*, 9:51–58.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Annual Statistics*, 7:1–26.

Efron, B. & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. Chapman & Hall.

Engle, R. F. (1984). Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics. In Griliches, Z. & Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, pages 776–825. North–Holland.

Engle, R. F. & Granger, C. W. J. (1987). Co-Integration and Error Correction: Representation, Estimation and Testing. *Econometrica*, 55:251–276.

Engle, R. F. & Yoo, B. S. (1987). Forecasting and Testing in Co–Integrated Systems. *Journal of Econometrics*, 35:143–159.

Feige, E. L. & Pearce, D. K. (1979). The Casual Causal Relationship Between Money and Income: Some Caveats for Time Series Analysis. *Review of Economics and Statistics*, 61:521–533.

Feigl, H. (1953). Notes on Causality. In Feigl, H. & Brodbeck, M., editors, *Readings in the Philosophy of Science*. Appleton–Century–Crofts.

Florens, J. P. & Fougère, D. (1996). Non-Causality in Continuous Time. *Econometrica*, 64:1195–1212.

Florens, J. P. & Mouchart, M. (1982). A Note on Noncausality. *Econometrica*, 50:583–591.

Florens, J. P. & Mouchart, M. (1985). A Linear Theory for Noncausality. *Econometrica*, 53:157–175.

Gaffke, N., Steyer, R. & von Davier, A. A. (1999). On the Asymptotic Null–Distribution of the Wald Statistic at Singular Parameter Points. *Statistics & Decisions*, 17:339–358.

Gallant, A. R. (1977). Testing a Nonlinear Regression Specification: A Nonregular Case. *Journal of the American Statistical Association*, 72:523–530.

Gallant, A. R. & Tauchen, G. (1989). Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications. *Econometrica*, 57:1091–1120.

Geweke, J. (1982). Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association*, 77:304–313.

Geweke, J. (1984). Inference and Causality in Economic Time Series Models. In Griliches, Z. & Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, pages 1102–1142. North–Holland.

Giles, J. A. (2000). Testing for Two–step Granger Noncausality in Trivariate VAR Models. Econometrics Working Paper 0008, University of Victoria, Canada.

Gleason, J. R. (1988). Algorithms for Balanced Bootstrap Simulations. *The American Statistician*, 42:263–266.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37:424–438.

Granger, C. W. J. (1980). Testing for Causality - A Personal Viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352.

Granger, C. W. J. (1986). Developments in the Study of Cointegrated Economic Variables. *Oxford Bulletin of Economics and Statistics*, 48:213–228.

Granger, C. W. J. (1987). Equilibrium, Causality and Error-Correction Models. *Economic Notes*, 1:5–21.

Granger, C. W. J. (1988). Some Recent Developments in a Concept of Causality. *Journal of Econometrics*, 39:199–211.

Granger, C. W. J., King, M. L. & White, H. (1995). Comments on Testing Economic Theories and the Use of Model Selection Criteria. *Journal of Econometrics*, 67:173–187.

Granger, C. W. J. & Lin, J.-L. (1995). Causality in the Long Run. *Economic Theory*, 11:530–536.

Granger, C. W. J. & Newbold, P. (1986). *Forecasting Economic Time Series*, pages 120–135 and 216–234. Academic Press Inc.

Gregory, A. W. & Veall, M. R. (1985). Formulating Wald Tests of Nonlinear Restrictions. *Econometrica*, 53:1465–1467.

Gupta, S. (1987). Testing Causality: Some Caveats and a Suggestion. *International Journal of Forecasting*, 3:195–209.

Hall, P. (1990). On the Relative Performance of Bootstrap and Edgeworth Approximations of a Distribution Function. *Journal of Multivariate Analysis*, 35:108–129.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer.

Hall, P. & Wilson, S. R. (1991). Two Guidelines for Bootstrap Hypothesis Testing. *Biometrics*, 47:757–762.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Härdle, W., Klinke, S. & Müller, M. (1999). *XploRe – Learning Guide*. Springer–Verlag.

Hatanaka, M. (1996). *Time–Series–Based Econometrics. Unit Roots and Co–Integrations*. Oxford University Press.

Hillmer, M. (1992). *Kausalanalyse makroökonomischer Zusammenhänge mit latenten Variablen*. Physica.

Hinkley, D. V. (1989). Bootstrap Significance Tests. *Bulletin of the International Statistical Institute*, 53.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81:945–970.

Hosoya, Y. (1977). On the Granger Condition for Non-Causality. *Econometrica*, 45:1735–1736.

Hosoya, Y. (1991). The Decomposition and Measurement of the Interdependence Between Second–Order Stationary Processes. *Probability Theory and Related Fields*, 88:429–444.

Hsiao, C. (1979). Autoregressive Modeling of Canadian Money and Income Data. *Journal of the American Statistical Association*, 74:553–560.

Hsiao, C. (1982). Autoregressive Modeling and Causal Ordering of Economic Variables. *Journal of Economic Dynamics and Control*, 4:243–259.

Hylleberg, S. & Mizon, G. E. (1989). Cointegration Mechanisms. *Economic Journal*, 99:113–125.

Jeong, J. & Maddala, G. S. (1993). A Perspective on Application of Bootstrap Methods in Econometrics. In *Handbook of Statistics*, volume 11, pages 573–610. North–Holland.

Johansen, S. (1988). Statistical Analysis of Cointegrating Vectors. *Journal of Economic Dynamics and Control*, 12:231–254.

Johansen, S. (1991). Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. *Econometrica*, pages 1551–1580.

Johansen, S. (1995). *Likelihood–Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.

Johansen, S. & Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration - with an Application to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52:169–210.

Johnston, J. & DiNardo, J. (1997). *Econometric Methods*. McGraw–Hill.

Judge, G. G., Hill, R. C., Griffiths, W. E., Lütkepohl, H. & Lee, T.-C. (1988). *Introduction to the Theory and Practice of Econometrics*. Wiley.

Kaufmann, S. (1996). Permanent Components in Swiss Macroeconomic Variables. *Schweizerische Zeitschrift für Volkswirtschaft und Statistik*, 132:539–561.

King, R. G., Plosser, C. J., Stock, J. H. & Watson, M. W. (1991). Stochastic Trends and Economic Fluctuations. *The American Economic Review*, 81:819–840.

Kirchgässner, G. (1981). *Einige neuere statistische Verfahren zur Erfassung kausaler Beziehungen zwischen Variablen*. Vandenhoeck & Ruprecht.

Kirchgässner, G. & Wolters, J. (1992). Implications of Temporal Aggregation on the Relation Between Two Time Series. *Statistical Papers*, 33:1–19.

Konishi, T. & Granger, C. W. J. (1992). Separation in Cointegrated Systems. Discussion Paper 51, University of California, USA.

Krämer, W. & Sonnberger, H. (1986). *The Linear Regression Model under Test*. Physica.

Lafontaine, F. & White, K. J. (1986). Obtaining Any Wald Statistic You Want. *Economics Letters*, 21:35–40.

Leamer, E. E. (1985). Vector Autoregressions for Causal Inference? In *Carnegie–Rochester Conference Series on Public Policy*, volume 22, pages 255–304.

Lütkepohl, H. (1982). Non–Causality Due to Omitted Variables. *Journal of Econometrics*, 19:367–378.

Lütkepohl, H. (1986). Comparison of Predictors for Temporally and Contemporaneously Aggregated Time Series. *International Journal of Forecasting*, 2:461–475.

Lütkepohl, H. (1989a). The Stability Assumption in Tests of Causality Between Money and Income. In Krämer, W., editor, *Econometrics of Structural Change*, pages 75–86. Physica.

Lütkepohl, H. (1989b). Prediction of Temporally Aggregated Systems Involving Both Stock and Flow Variables. *Statistical Papers*, 30:279–293.

Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Springer.

Lütkepohl, H. (1993). Testing for Causation Between Two Variables in Higher–Dimensional VAR Models. In Schneeweiß, H. & Zimmermann, K., editors, *Studies in Applied Econometrics*, pages 75–91. Physica.

Lütkepohl, H. (1994a). Problems Related to Testing for Granger-Causality in VARMA Processes. SFB 373 Discussion Paper 9, Humboldt Universität, Berlin, Germany.

Lütkepohl, H. (1994b). Kointegration und gemeinsame Trends. SFB 373 Discussion Paper 28, Humboldt Universität, Berlin, Germany.

Lütkepohl, H. (1994c). Interpretation of Cointegration Relations – Comments on "Estimating Systems of Trending Variables" by S. Johansen. *Econometric Reviews*, 13:391–394.

Lütkepohl, H. (1996a). *Handbook of Matrices*. John Wiley & Sons.

Lütkepohl, H. (1996b). Testing for Nonzero Impulse Responses in Vector Autoregressive Processes. *Journal of Statistical Planning and Inference*, 50:1–20.

Lütkepohl, H. (1999). Vector Autoregressive Analysis. SFB 373 Discussion Paper 31, Humboldt Universität, Berlin, Germany.

Lütkepohl, H. & Breitung, J. (1997). Impulse Response Analysis of Vector Autoregressive Processes. In Heij, C., editor, *System Dynamics in Economics and Financial Models*, pages 299–326. John Wiley & Sons.

Lütkepohl, H. & Burda, M. M. (1997). Modified Wald Tests under Nonregular Conditions. *Journal of Econometrics*, 78:315–332.

Lütkepohl, H. & Poskitt, D. S. (1996). Testing for Causation using Infinite Order Vector Autoregressive Processes. *Econometric Theory*, 12:61–87.

Lütkepohl, H. & Reimers, H.-E. (1992a). Impulse Response Analysis of Cointegrated Systems. *Journal of Economic Dynamics and Control*, 16:53–78.

Lütkepohl, H. & Reimers, H.-E. (1992b). Granger-Causality in Cointegrated VAR Processes: The Case of the Term Structure. *Economics Letters*, 40:263–68.

Lütkepohl, H. & Saikkonen, P. (1997). Impulse Response Analysis in Infinite Order Cointegrated Vector Autoregressive Processes. *Journal of Econometrics*, 81:127–157.

MacKinnon, J. G. (1991). Critical Values for Co–Integration Tests. In Engle, R. F. & Granger, C. W. J., editors, *Long–Run Economic Relationships*, pages 267–276. Oxford University Press.

Maddala, G. S. & Kim, I. (1998). *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.

Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*. Springer.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1974). *Multivariate Analysis*. Academic Press.

Mellander, E., Vredin, A. & Warne, A. (1992). Stochastic Trends and Economic Fluctuations in a Small Open Economy. *Journal of Applied Econometrics*, 7:369–394.

Mosconi, R. & Giannini, C. (1992). Non-Causality in Cointegrated Systems: Representation, Estimation and Testing. *Oxford Bulletin of Economics and Statistics*, 54:399–417.

Newbold, P. (1982). Causality Testing in Econometrics. In Anderson, O. D., editor, *Time Series Analysis: Theory and Practice I*, pages 701–716. North–Holland.

Osterberg, W. P. (1992). Inflation, Financial Structure, and the q Theory of Investment: An Empirical Analysis of Short–Run and Long–Run Causal Relations. *Research in Finance: A Research Annual*, 10:129–149.

Park, J. & Phillips, P. C. B. (1989). Statistical Inference in Regressions with Integrated Processes: Part 2. *Econometric Theory*, 5:95–131.

Paulsen, J. (1984). Order Determination of Multivariate Autoregressive Time Series with Unit Roots. *Journal of Time Series Analysis*, 5:115–127.

Penm, J. H. W. & Terrell, R. D. (1982). On the Recursive Fitting of Subset Autoregressions. *Journal of Time Series Analysis*, 5:115–127.

Penm, J. H. W. & Terrell, R. D. (1984). Multivariate Subset Autoregressive Modelling with Zero Constraints for Detecting "Overall Causality". *Journal of Econometrics*, 23:311–330.

Pesaran, M. H. & Shin, Y. (1998). Generalized Impulse Response Analysis in Linear Multivariate Models. *Economics Letters*, 58:17–29.

Phillips, P. C. B. (1995). Fully Modified Least Squares and Vector Autoregression. *Econometrica*, 63:1023–1078.

Phillips, P. C. B. & Hansen, B. E. (1990). Statistical Inference in Instrumental Variables Regressions with I(1) Processes. *Review of Economic Studies*, 57:99–125.

Pierce, D. A. (1975). Forecasting in Dynamic Models with Stochastic Regressors. *Journal of Econometrics*, 3:349–374.

Pierce, D. A. & Haugh, L. D. (1977). Causality in Temporal Systems - Characterizations and a Survey. *Journal of Econometrics*, 5:265–293.

Reinsel, G. C. & Ahn, S. K. (1992). Vector Autoregressive Models with Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test, and Forecasting. *Journal of Time Series Analysis*, 13:353–375.

Renault, E., Sekkat, K. & Szafarz, A. (1998). Testing for Spurious Causality in Exchange Rates. *Journal of Empirical Finance*, 5:47–66.

Renault, E. & Szafarz, A. (1991). True Versus Spurious Instantaneous Causality. Working paper, GREMAQ and IDEI, Toulouse, France.

Savin, N. E. (1984). Multiple Hypothesis Testing. In Griliches, Z. & Intriligator, M. D., editors, *Handbook of Econometrics*, volume 2, pages 828–879. North–Holland.

Schneider, W. (1991). Kausale Strukturen: a priori Restriktionen oder testbare Eigenschaften ökonomischer Modelle? Working paper, Christian–Albrechts Universität, Kiel, Germany.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley.

Sheehan, R. G. & Grieves, R. (1982). Sunspots and Cycles. A Test of Causation. *Southern Economic Journal*, 49.

Sims, C. A. (1972). Money, Income and Causality. *The American Economic Review*, 62:540–552.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48:1–48.

Sims, C. A. (1981). An Autoregressive Index Model for the U.S., 1948-1975. In Kmenta, J. & Ramsey, J. B., editors, *Large–Scale Macro–Econometric Models*, pages 283–309. North–Holland.

Sims, C. A., Stock, J. H. & Watson, M. W. (1990). Inference in Linear Time Series Models with Some Unit Roots. *Econometrica*, 58:113–144.

Stock, J. H. & Watson, M. W. (1988). Testing for Common Trends. *Journal of the American Statistical Association*, 83:1097–1107.

Stock, J. H. & Watson, M. W. (1989). Interpreting the Evidence on Money-Income Causality. *Journal of Econometrics*, 40:161–181.

Swanson, N. R. & Granger, C. W. J. (1992). Impulse Response Functions Based on a Causal Approach to Residual Orthogonalization in Vector Autoregressions. *Journal of the American Statistical Association*, 92:357–367.

Tiede, M. (1991). Unsichere statistische Befunde von Granger–Kausalitätsbeziehungen. *Jahrbücher für Nationalökonomie und Statistik*, 208:98–102.

Tobin, J. (1970). Money and Income. Post Hoc Ergo Propter Hoc? *The Quarterly Journal of Economics*, 84:301–317.

Toda, H. Y. (1991). *Vector Autoregressions and Causality*. PhD thesis, Yale University, USA.

Toda, H. Y. & Phillips, P. C. B. (1993). Vector Autoregressions and Causality. *Econometrica*, 61:1367–1393.

Toda, H. Y. & Phillips, P. C. B. (1994). Vector Autoregressions and Causality: a Theoretical Overview and Simulation Study. *Econometric Reviews*, 13:259–285.

Toda, H. Y. & Yamamoto, T. (1995). Statistical Inference in Vector Autoregressions with Possibly Integrated Processes. *Journal of Econometrics*, 66:225–250.

Triacca, U. (1998). Non–Causality: the Role of the Omitted Variables. *Economics Letters*, 60:317–320.

Triacca, U. (2000). On the Hsiao Definition of Non–Causality. *Economics Letters*, 66:261–264.

Vercelli, A. (1992). Causality and Economic Analysis: A Survey. In Vercelli, A. & Dimitri, N., editors, *Macroeconomics: a Survey of Research Strategies*, pages 393–421. Oxford University Press.

Vuong, Q. H. (1987). Generalized Inverses and Asymptotic Properties of Wald Tests. *Economics Letters*, 24:343–347.

Warne, A. (1990). *Vector Autoregressions and Common Trends in Macro and Financial Economics*. PhD thesis, Stockholm School of Economics, Stockholm, Sweden.

Warne, A. (1992a). Wald Tests of Nonlinear Restrictions in Vector Autoregressions with Integrated Time Series. Working paper, Institute for International Economic Studies, Stockholm University, Sweden.

Warne, A. (1992b). Inference in Vector Autoregressions with Integrated Time Series. Working paper, Institute for International Economic Studies, Stockholm University, Sweden.

Warne, A. (1997). Inference in Cointegrated VAR Systems. *The Review of Economics and Statistics*, 79:508–511.

Wiener, N. (1956). The Theory of Prediction. In Beckenback, E. F., editor, *Modern Mathematics for Engineers*, 1, chapter 8. McGraw–Hill.

Wolters, J. (1991). Alternative Ansätze zur Messung der Persistenz in ökonomischen Zeitreihen. *Allgemeines Statistisches Archiv*, 75:184–196.

Yamada, H. & Toda, H. Y. (1997). A Note on Hypothesis Testing Based on the Fully Modified Vector Autoregression. *Economics Letters*, 56:27–39.

Yamada, H. & Toda, H. Y. (1998). Inference in Possibly Integrated Vector Autoregressive Models: Some Finite Sample Evidence. *Journal of Econometrics*, 86:55–95.

Zapata, H. O. & Rambaldi, A. N. (1997). Monte Carlo Evidence on Cointegration and Causation. *Oxford Bulletin of Economics and Statistics*, 59:285–298.

Zellner, A. (1979). Causality and Econometrics. In *Carnegie Rochester Conference Series on Public Policy*, volume 10, pages 9–54.

Curriculum

Name:	Maike M. Burda
10/1986 - 09/1991	Studium an der Christian–Albrechts–Universität zu Kiel, Fachrichtung Volkswirtschaft (quantitative Richtung).
10/1991 - 9/1992	Wissenschaftliche Mitarbeiterin an der Christian–Albrechts–Universität zu Kiel, Lehrstuhl Prof. Dr. Gerd Hansen, Institut für Statistik und Ökonometrie.
10/1992 - 9/1999	Wissenschaftliche Mitarbeiterin an der Humboldt–Universität zu Berlin, Lehrstuhl Prof. Dr. Helmut Lütkepohl, Institut für Statistik und Ökonometrie.

Selbständigkeitserklärung

Hiermit erkläre ich, die vorliegende Arbeit selbständig ohne fremde Hilfe verfasst zu haben und nur die angegebene Literatur und Hilfsmittel verwendet zu haben.

Maike M. Burda

28. August 2001